A Computational Approach to Gestural Interactions of the Upper Limb on Planar Surfaces

Antoine Loriette

Submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy

School of Computing Science

College of Science and Engineering
University of Glasgow

March 2019

Abstract

There are many compelling reasons for proposing new gestural interactions: one might want to use a novel sensor that affords access to data that couldn't be previously captured, or transpose a well-known task into a different unexplored scenario. The *creation, optimisation or understanding* of new interactions remains, however, a challenge. Models have been used to foresee interaction properties: Fitts' law, for example, accurately predicts movement time in pointing and steering tasks. But what happens when no existing models apply?

This thesis, carried out within a context of design for users with reduced mobility, proposes to investigate new gestural interactions around planar surfaces involving the upper limb only, which are the results of a dialogue and design workshops with occupational therapists. For instance, the task of text-input is ported to an interaction of the upper limb through optically-tracked virtual surfaces, the activity of arm reach rehabilitation is associated to digital gameplay for improved user engagement and the elicitation of motions users can produce is automated through the production of audio rewards.

The core assertion to this work is that a computational approach provides the frameworks and associated tools that are needed to model such interactions. In addition, it offers solutions to address the challenges the creation, optimisation or understanding such interactions present. We support this assertion through three research projects.

In Chapter 3, a closed-loop model of the interaction is used in which sensor outputs of a RGB-D camera are transformed through signal processing into variables that can drive a gesture typing interaction on off-the-shelf mobile devices. The need for regression and classification of processed variables is addressed by using a Kalman filter and a neural network. In Chapter 4, the interaction is cast as an optimisation problem which should balance a dual objective of arm reach rehabilitation and user engagement. Searching for design configuration requires a low-latency measure of reference gameplay, which is constructed as a probabilistic model of variables captured in both game space and game controller space. In Chapter 5, a model of user motions, based on the concatenation of sensor outputs and their derivatives, provides access to quantifiable properties: variability and repeatability. By using a model of expected user motions, a discriminative model can be trained with synthetic data to classify the repetitive nature of unseen user motions.

In all chapters, users studies are carried out to measure the influence of different parameters on interactions properties. Effects of scale, in particular, are investigated. Finally, post hoc data analyses are undertaken to shed light on effects and behaviours observed with real users.

Acknowledgements

I would like to thank my supervisors John Williamson and Roderick Murray-Smith for the help and inspiration they provided me with during the four years of research this thesis took to complete. Thanks also to my closest collaborators, Sebastian Stein and Andrew Ramsay, who showed great support on a daily basis. The IDI group and the Moregrasp consortium have always supplied an environment suitable for insightful discussions and supportive comments; Bjorn Jensen, Catherine Higham and Simon Rogers in particular.

My work has benefited from precious interactions with practioners involved with the Queen Elisabeth Univeristy of Glasgow, occupational therapists from its spinal unit and the members of the Spinal Injury Scotland charity. I am thankful to Leslie Wallace, Michelle Rankin, Jennifer Cloughley, Emily Houston, Louise Cownie, Margaret Purcell and Emma-Jane Gault for their time, their ideas and feedback as well as their involvement in a series of workshops.

I wish also to thank the many other students with whom I have spent my time in the office and which company has kept me happy and entertained: Sven, Dominik, Daniel, Lauren, Daniel, Xiaoyu, Miroslav, Iulia, Natasha, Grymur, with a special mention to Anders and Francesco who also supported me outside of the School.

The whole idea of getting engaged with doing research has some older roots. Barry, Donny, Moira, Kia and the Pedros have inspired me to go through this infamous journey; while Mathias allowed me to look for new projects in the first place.

To the friends I made on the way, Tristan, Ayley, Pierre, Dong, Meredith and Sarah. Thanks. And to my long time friends Etienne, Malec, Julie, Camille, Vicent, Severine, Sebastian, Kahina and the Randomers, thanks for the time we spent, always.

Lastly, a lot of gratitude goes towards my family, my parents Philippe and Aline, my brother Sylvain, sisters Camille and Bérénice, to their associated offsprings, as well as my grandparents. You have been providing me with a lot happy moments during this endeavour. Merci pour tout.

I realise the categories above are not all well organised, and I must also have forgotten many of you. Sorry about that.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

There are many compelling reasons for proposing new gestural interactions. When a new sensor is released, new types of data are often made available, which can in turn be used for interaction. For instance, the Microsoft Kinect sensor, exposing depth data and body pose data, spawned a new set of interactions leveraging these previously inaccessible sources of information. This sensor has been used to make inert surfaces responsive to touch interactions [1] or help users learn physical movement sequences [2]. Similarly, the inclusion of additional body parts has also created opportunities for new interactions. Making use of the controllable degrees of freedom in rotation afforded by the hand holding a mobile device, new techniques for map navigation have been proposed [3] and by equipping a user shoe with a camera, new gestural input techniques have been investigated [4]. Aside from new sensors or new body parts, new tasks, such as gesture typing [5], and new scenarios can also inspire new gestural interactions, affording users with novel and improved experiences. In this thesis, specifically, a subgroup of gestural interactions is focused on: those recruiting the upper limb only and situated around planar surfaces.

Upper limb gestural interactions on planar surfaces are defined as interactions in which the arm of the user is motioning across and in contact with a plane, such as a tabletop, and for which the need for fine motor control of the fingers does not necessarily exist. Traditional interactions, such as the one afforded by a mouse computer, belong to this category, as well as interactions based on large touch sensitive displays found in public areas. Usual ways of proposing such interactions have involved dedicated hardware to create the enabling surfaces. Recently, new exciting scenarios have been proposed thanks to advances in optical tracking. Cameras have thus been used to track hand positions and classify the touching behaviour of fingers over arbitrary surfaces, potentially affording the same kind of interactions as large touch sensitive screen with the added benefit of providing mobility and sensing over the surface, opportunistically extending the interaction beyond touch and towards mid-air interactions [6].

Upper limb interactions usually involve large dimensional areas. These have been employed for presenting graphical information, potentially reducing visual clutter, or enabling multi-users experiences with space sharing. They have as such demonstrated a high versatility of usage with applications in the field of data visualisation, computer-assisted collaborative work or rehabilitation. Given the benefits granted by large surfaces and dimensions, these interactions often recruit more muscle groups than other interactive systems such as mobile phones or desk computers, and in addition, the range of motions needed for interaction do not necessarily compare with more stationary interaction scenarios, both factors highlighting the prevalence of the user body in the realisation of the interaction. For example, different regions have been identified on tabletops with easily reachable areas being described as more personal than farther regions which extend towards neighbouring users or require users to lean over the surface [7]. In mouse interaction, clutching, the act of lifting the device, is a common strategy and arises when users reach their limits in terms of arm motion range. In other words, upper limb gestural interactions on planar surfaces make apparent that effects of scale and limits of the user body need to be considered for their understanding.

After an initial design phase, a new upper limb gestural interactions can be considered from a computational perspective where an interaction model is proposed. This is apparent at different levels. First, the interaction needs to be implemented or created. Models are commonly used for this task, where output of sensors are transformed into controllable variables and system dynamics are set. The complexity of this processing step depends on the variety and type of input data available. Extracting information from an image or a motion requires indeed more work than from the binary output of a push-button. Second, the interaction can benefit from the optimisation of its parameters. It is not always possible to predict what configurations will favour user performance or preference. Effects of scale, for instance, should have an impact which is seldom modelled. As a result, users studies are often employed to map the relationships between design variables and task outcomes, but models can here also be used to find such configurations. Finally, as the new interaction materialises, a previously unexplored space can be studied. New effects might be uncovered, and data samples from recorded interactions can be analysed. Fitts' law is again here a good example, where properties emerging from the data were summarized into a powerful predictive model. Ideally, the lessons learned at each of these stages should be fed back into the design process, enabling iterative improvements to take place.

Through different case studies, this thesis is aiming at modelling new upper limb gestural interactions on planar surfaces. It follows a computational approach to design, where models will be employed to create, optimise or understand interactions.

1.1 Thesis Statement

There are many compelling reasons for proposing new upper limb gestural interactions. But the lack of effective models presents a challenge for their creation, optimisation or understanding. The main assertion to this work is that a computational approach to design provides the modelling tools needed to address each of these challenges.

Novel gestural upper limb interactions are approached from a computational perspective, where discriminative models are used to enable interactions, optimisation is included as an integral part of their design and reinforcement learning is used to explore motions users produce in such interactions.

1.2 Research Contributions

The contributions made by this thesis are:

In Chapter 3:

- A system that affords text-input through gesture typing on planar surfaces created by means of visual tracking.
- A discriminative model for the classification of contact between fingertip and planar surfaces from depth image data.
- A study of the influence of scale on text-input performance and on user preference.

In Chapter 4:

- A system that affords play in rehabilitation of users with limited mobility, while enacting motions recommended by therapists.
- A measure of user performance suitable for optimisation, relatively user independent, which is based on the probabilistic modelling of user input and game state.
- A formulation of the use of digital games in physical rehabilitation as a dual optimisation problem between rehabilitation goals and user engagement.

In Chapter 5:

- An automated elicitation process of user repetitive motions.
- A model for detecting and segmenting repetitive motions in real-time sensor data.
- A quantitative measure of unconstrained user motions for upper limb gestural interactions on planar surfaces.

1.3 Overview of the Thesis

The first chapter provides background information about users having sustained a spinal cord injury. It summarises work related to upper limb interactions around planar surfaces and gestural interactions in general. Common models, as well as modelling techniques, used in the field of Human and Computer Interactions (HCI) are finally presented.

It is followed by three research sections, each constituting a single chapter:

- **Gesture Typing through Virtual Tabletops** (Chapter 3):
  proposes the design of gesture typing through virtual surfaces on tabletops. The problem of fingertip touch classification is solved with a supervised learning approach and the measure of writing performance against capacitive touch systems and as a function of scale is conducted through a user study.

- **Rehabilitation through Common Gameplay** (Chapter 4):
  investigates an alternative method to foster user motivation towards rehabilitation exercises by using ready-made games and adaptive controls. A probabilistic model of user behaviour is built from a dataset of reference play and is used to relate user actions and in-game variables to game score.

- **Eliciting Motions Through Positive Reinforcement** (Chapter 5):
  looks into modelling motions through the definition of a joint user-sensor space. The user ability to produce variability and reliability in their motions is investigated. A supervised learning approach is used to identify and segment repetitions.

Due to the heterogeneous nature of the focus of each research section (text-input, physical rehabilitation and elicitation of motions), each chapter also includes a topical literature review at its beginning or before new concepts are introduced.

The final chapter summarises the thesis' research contributions and results, discusses the limitations to this work and concludes with some avenues for further research.

Declaration of Originality

I confirm that the material presented here is the result of my own work without collaboration.

Chapter 2

Background

2.1 Context

Defining a subset of HCI interactions by removing the need for fine motor control for the fingers might seem unnatural at first glance: grasping through the presence of an opposed thumb is considered as the most important movement in the human hand [8] and, as a complex and diverse function, is usually required in the everyday interactions we have with computing devices. Other hand functions exist however, non-prehensile movement such as pushing, lifting, tapping or punching do not always necessitate fine finger motor control and cover many interactions already available or proposed in the field of HCI. For instance, upper limb interactions include the most prevalent smartphone's single point-of-touch interaction but excludes any multi-touch interactions where complex inter-finger coordination comes in play. Using intermediate joints such as the wrist through tilt [3] and rotation gestures [9] also falls into this categorisation. A good illustration of the modularity of upper limb interactions can be found in the focused-casual continuum. Pohl et al. [10] proposed a task in which a user is allowed three interaction techniques to steer a cursor through a tunnel: touch, hover and gestures. Users adapted their behaviours to the required level of control, but choose the loosest one when not given the choice. While diverse in terms of control they provide and sensing they require, none of these techniques required fine finger motor control and can be described solely as upper limb interactions.

Despite the de facto existence and interesting properties of the upper limb interaction category, the main reason behind its definition resides in the presence of a user group for which fine motor control of end-effectors has been compromised by an injury to the spinal cord. The European project Moregrasp[1] [11], which has funded this research, was dedicated to improving the life of users with spinal cord injury through the use of a neuroprosthesis. The aim of

[1] http://www.moregrasp.eu/

the MoreGrasp project was to develop a "non-invasive, multi-adaptive, multimodal user interface including a brain-computer interface (BCI) for intuitive control of a semi-autonomous motor and sensory grasp neuroprosthesis supporting individuals with high spinal cord injury in everyday activities". This project was organised as a partnership between different stakeholders ranging from health practitioners, medical engineers, computing scientists and users who had sustained a spinal cord injury, creating a rich environment for the understanding of how technology could help mitigate the severe consequences of motor impairments. Given the specific motor limitations of the main target user group of the Moregrasp project, upper limb gestural interactions emerged as a suitable subcategory for research in this thesis.

Spinal cord injury

In Europe, 11,000 new cases of spinal cord injuries (SCI) are registered each year bringing the total current population of injured persons to about 330,000. The level of the injury distinguishes two types of impairments: paraplegia and tetraplegia, with the latter meaning that not only the lower but also the upper extremities are paralysed. Tetraplegia is the most common impairment with more than half of the new injuries. Injuries have different origins and are split between traumatic, most likely to be the consequence of a fall, a road accident or the practise of sport and leisure activities, and non-traumatic usually caused by tumours (benign and malignant), inflammatory, vascular or degenerative causes. A bigger proportion of male is represented in the population with traumatic SCI, while the population with non-traumatic SCI is balanced in terms of gender. Studies [12, 13] surveying the priorities in term of functional recoveries for the spinal cord injured population have showed a strong dichotomy between paraplegic and tetraplegic respondents. While paraplegic respondents signalled sexual function with the highest priority, tetraplegic respondents showed a strong focus on the upper limb with arm and hand function recovery designated as the main factor for improving their quality of life (Figure 2.1, right). To understand this dichotomy, two pieces of information are needed. The map of key sensory points by Maynard et al. [14], reproduced on Figure 2.1 (left), illustrates the procedure for identifying the level of a SCI through probing for the presence or lack of sensation on the patient's skin. Side to side with the vertebrae nomenclature, adapted from [15], these provide an idea of the connection between the level of an injury and its impact on the anatomical level. For low levels of injury, up to lumbar vertebrae, only the legs of the patients are impacted. For very high levels of injury, typically in the neck or above the shoulders, the nerves sending signals to the forearms muscles that control finger motions have been partially or completely severed and can not any longer fulfil their intended functional role.

Users with injuries at level C5 and C6 were the target group of the Moregrasp project. This motivates restricting gestural interaction to the ones only recruiting the upper limb and not requiring fine motor control of the fingers. Additionally, the help of a surface to rest ones arm has also been identified as a requirement in order to limit the reliance on shoulder muscles.

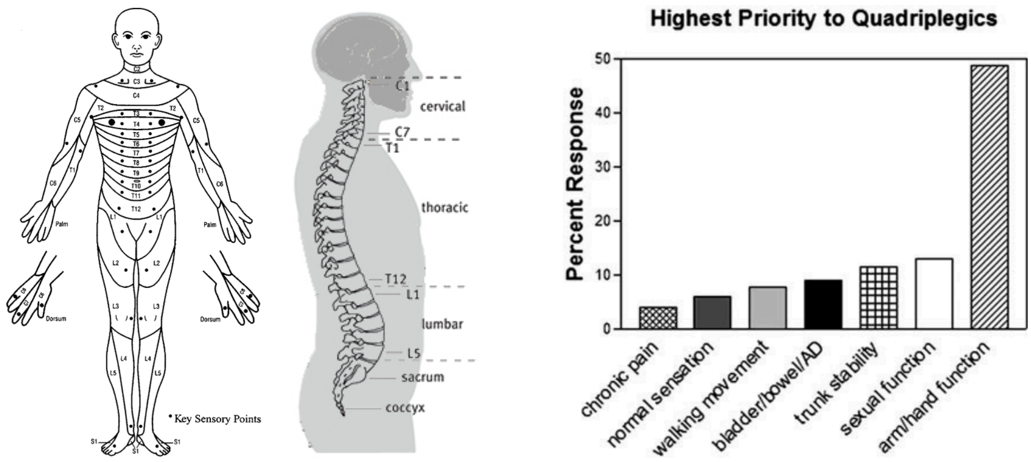

Figure 2.1: Information linking the level of spinal cord injury to the extent of loss of sensory feeling on the body. Keypoints for assessing the level of SCI [Maynard - 1997] (left). Vertebra nomenclature, adapted from [Kayalioglu - 2009] (middle). Highest identified priorities for tetraplegic users [Anderson - 2004] (right).

HCI And SCI

Users with SCI are usually not the main target in studies related to accessibility, but are often part of a wider group presenting similar motor impairments such as those caused by strokes, tremors or Parkinson's disease. From an HCI perspective, mobile interaction has become a focal point due to the prevalence of mobile phones as principal computing and communication devices. As such, studies centred around mobile interactions provide a commanding viewpoint into how people interact with technology. There exists however very specific research projects which focus has involved interactive planar surfaces. Here, a selective review of research explicitly related to users with motor impairment, touch interaction and tabletop rehabilitation is presented.

Beyond the intent behind mobile information needs [16, 17], which can be assumed to be independent of the presence or not of motor disability, interaction techniques, principally touch as a consequence of its prevalence, have been the focus of studies with users with motor disabilities. In term of interaction, touch shines as a complex task. Qualitative research, with an analysis of user generated videos from a popular social website by Anthony et al. [18], showed that users with motor disabilities recruited different body parts, the index finger predominantly but also their thumb, knuckles, hand or even nose, to interact with a smartphone touchscreen and overcome physical limitations. Multitouch interaction has been proven to be difficult for a majority of users and the issue with spurious touches, in some cases called "palm rejection problem" is pointed out for some users did not managed to restrict their interaction with the touch sensitive surface to only one point of contact. Another qualitative study by Naftali et al. [19], with an online survey, a diary analysis and contextual sessions, has shown that text input remains the hardest task both at home and on-the-go, only seconded by text correction. They also point out that mobile phones are often used while being supported by another surface (user lap or tabletop); an observation also present in [18].

On a more quantitative level, measuring touchscreen performance for users with motor disabilities has been the topic of several studies. Montague et al. [20] instrumented a Sudoku game to capture users touch location and intent, providing the information needed for deriving successes and errors during the interaction. They showed that user performance is highly variable between participants and between participant's session themselves. With a focus on users with SCI only, Guerreiro et al. [21] investigated the performance of different input techniques including tapping and crossing target as well as linear dragging. In particular, they found out that the performance of directional gesturing was highly dependent on the direction of the gesture with movement along the screen's diagonal being the most prevalent to errors. Finally, Findlater et al. [22] were interested in the comparing mouse and touchscreen performance for users with upper body motor impairment and included a control group of users without motor impairment. They showed that touch was faster than mouse but more error prone, that tapping was three times more prone to errors for users with impairment and that again spurious touches were problematic.

These research studies reveal that while touch interaction on mobile is a complex task, multi-touch interaction poses even greater challenges to users. Spurious touches is regularly identified as an issue for users with motor impairments. The same is true for input errors, in tapping in particular. The findings from this selective literature will be used later in the thesis as a seed for the design of the interaction presented in Chapter 3.

Aside to these touch scenarios, interactive tabletops have also been used with older users and users with rehabilitation needs. With a qualitative study involving older adults, Piper et al. [23] explored the accessibility and appeal of surface computing. They reported that with surface computing, just like touch, some gestures that required two fingers or fine motor movement were problematic. Ratings for ease of use and ease of performing each action as well as time required to figure out an action were similar to that of younger adults. Older adults reported that the surface computer was less intimidating, less frustrating, and less overwhelming than a traditional computer. The idea of using a surface computer for health care support was well-received by participants. On an application level, Augstein et al. [24, 25] with their tabletop rehabilitation project have explored different games for the cognitive training of older adults, and also studied how health practitioners were supported in their rehabilitation work by technology. They reported that the main two complaints for the therapists were focused on the technical limitations of the system and the lack of diversity in the games proposed.

The diversity in motor disabilities users exhibited in the studies aforementioned shows that there is more than just a binary picture when it come to impairments. There exists a continuum from non-impaired users to impaired users, and similarities between situationally impairment users, who will for example be temporarily restricted in their movement capabilities, and more permanent motor disabilities. Designing for disabilities has provided some great opportunities. Pullin [26] demonstrated how an innovative approach transformed the way eye glasses were perceived and provided benefits for a majority. Challenges remain however as assistive technologies sometimes mark their users as having disabilities, Shinohara et al. [27].

2.2 Upper limb Interactions on Planar Surfaces

Upper limb interaction on planar surfaces are best represented in the literature on interactive tabletops and surface computing. Such designs have usually tried to bring the virtual world of computers onto the desk that supports them, bringing the desktop to the desk's top to make use of the extra real estate and connect virtual and physical manipulation of digital content. In 1993, Wellner et al. [28] proposed *Paperdesk*. In their paradigm, a conventional desk was instrumented with cameras and projectors enabling the projection of digital content onto the planar surface and any object lying over it. Users interacted with their bare hands with such digital content. Several use cases were envisioned, such as the interaction with a virtual calculator that users would simply touch to action or desk sharing with the representation of the actions of another remotely present users onto the projected graphical interface. No mentions of the upper limb are present in this paper which focuses on the users hand only, but reports from test users mention first that they had "more space" than a traditional work station, showing the importance of the physical dimension when users move from a desktop environment to a device supporting upper limb interactions. Other types of sensing have been proposed, such as *DiamondTouch* by Dietz et al. [29], where large surfaces where instrumented with capacitively coupled antennas able to detect touches from several users. More recently and with the help of new hardware, Hilliges et al. [30] have developed these initial ideas further. Interactive surfaces, that can sense whole hands and physical objects placed on them, were created using large touch sensitive tabletops. Models of interaction based on multi-touch and game physics simulations allowed tangible interactions with high fidelity with the real world.

From a sensing point of view, surface computing and interactive tabletops present the main drawback of requiring expensive and bulky contraptions that make their usability an issue in places where portability and adaptability are important.

Optical sensing has in that regards proposed some interesting solutions. Since the release of newly commoditised hardware such as depth cameras, opportunities for new interactions generated by optically tracked surfaces have arisen. These cameras provide a measure of distance between their sensor and any objects in line of sight, essentially adding a new dimension to every pixels of a conventional colour image. The acronym RGB-D, with D standing for depth, has been coined to reflect this idea. When the depth information is present, it becomes possible to understand the scene that is being captured: the distance and relationships between objects can be inferred and information as detailed as "a hand is hovering over the table with a finger in position (x,y,z)" can be extracted from the RGB-D video feed.

Leveraging the Kinect sensor from Microsoft, Wilson et al. [1] first demonstrated that depth cameras can be used as a touch sensor, opening the field to new experimentations. Harrison et al. [31] proposed to opportunistically create touch interfaces with associated video feedback on various surfaces such as the palm of the hand, the arm, a notepad or a wall, affording touch-like functionality comparable to that of a smartphone. More recently, Xiao et al. [32] focused their efforts on multitouch interaction over a large tabletop and provided a comparison for the tracking performance of existing algorithms, demonstrating an improvement both in term of tracking and touch detection. This was measured through target acquisition and steering tasks, both executed in an open-loop manner where no continuous feedback from the system was provided to the participant. Virtual reality have also benefited from optical tracking [33], showcasing the versatility of this approach.

From an interaction point of view, optical tracking offers some advantages. It could for example propose new solutions to the palm rejection problem which was identified for users with motor impairment. This problem currently need special processing such as the one proposed by Mott et al. [34] to be resolved. However, research with optically tracked surfaces has also consistently reported the need for better tracking performance which remains inferior to that of capacitive touchscreen. The design choices made in Chapter 3, which effectively set up an anchor point for the rest of this thesis, are grounded in the learnings from this selective literature review. For instance, the importance and difficulty of entering text through a mobile phone interaction for users with SCI has made text input a task of choice for the first experiment. The difficulty of performing tapping and the relative good performance of gesturing made apparent that gesture typing could be an interesting interaction technique. Finally, the recurrent problem with spurious touches, and the observation that mobile phones were interacted with on the user's lap or on a tabletop made a case for the use of optically tracked planar surfaces.

In summary, the context of the Moregrasp project was used as a seed for the design and the study of upper limb gestural interactions. I would argue that users with SCI act as a magnifying glass with interaction issues with technology that potentially pervade everybody's practise. As such, this thesis focuses on issues inspired by the impact of SCI on interactions but is aimed at a more general audience.

2.3 Gestural Interactions

There are several ways to characterise gestural interactions. On one hand, gestural interactions are simply interactions involving gestures, where gestures are defined as motions which are carrying information [35]. Gestures elicitation studies [36], which purpose is to identify user motions that seem to convey a recognisable intent, are a good illustration for the diversity in terms of types of motions [37, 38] that gestural interactions support. Pointing, drawing, tapping in rhythms [39], performing specific poses with hands or body have all been proposed. On the other hand, gestural interactions have also been partly characterised by the task of gesture recognition. Indeed, the capture of motions sometimes relies on sensors that do not always produce outputs readily usable for interaction, by opposition to push-buttons for example, and the task of gesture recognition becomes an integral part of their definition [17, 40]. It is interesting to note that pressing a button has also been viewed through this lens: Pohl et al. [41], even if they do not explicitly use the term gesture, used button presses to recognise users, casting this apparently simple task as gesture recognition. In the rest of this thesis, the definition used for gestural interaction is that of an interaction between two parties through a sensing device, involving any motions of the user body in order to convey some information.

The most prevalent gestural interaction, pointing, is also referred to as target acquisition. In the case of a traditional mouse and desktop interaction, pointing is the task of placing the device in control space so as its representation in display space is positioned at a desired location. The control space being the region where users interact, while the display space is the region where the effect of their actions is represented. The act of selection, signalled by an additional user action, completes the task of target acquisition.

This gestural interaction has been extensively studied through experiments where targets, of different sizes and placed at different distances from the users, are meant to be acquired. The user movement time for the interaction is measured, as the dependent variable, as a function of the size and distance of a given target, the independent variables. When users are instructed to be as fast and accurate as possible, such experiment is described as a Fitts' task and is deemed a spatially constrained task[2]. Fitts' law [43] is derived from such experiments, and has been used as a model for predicting the movement time. The Shannon formulation,

[2] A competing model has been proposed as the Linear Speed-Accuracy Tradeoff when users are instructed to aim for a given movement time, the task is then temporally constrained [42].

defined by MacKenzie et al. [42], relates the total movement time MT to the intended target's width W, lying at a distance D through the following equation:

$$MT = a + b \log_2 \left(\frac{D}{W} + 1 \right)$$

The constants a and b depend on the pointing technique and/or device being used and are usually measured through experimentation. An index of difficulty (ID) of the task, measured in bits, is defined by the logarithmic term:

$$ID = \log_2 \left(\frac{D}{W} + 1 \right)$$

The fraction between D and W encodes the idea that difficulty increases as a target moves farther away, or as a target becomes smaller. An index of performance (IP) can be defined as the ratio between index of difficulty and movement time:

$$IP = \frac{ID}{MT}$$

One of the advantage of such modelling is that the index of difficulty can be used to compare tasks, while the index of performance can be used to compare different input devices.

Other task have also been characterised in the same fashion. Typing on a keyboard, for example, can be viewed as a series of target acquisitions, and tasks which require users to follow trajectories, driving a pointer through a tunnel or as navigating a menu, have been shown to obey the "steering law" [44]. This led to the introduction of another type of gestural interaction created by composing trajectory-based tasks and text-input. Kristensson et al. [5] proposed *gesture typing* where users motion a pointer over the letters constituting the word they wish to input, and a decoding algorithm infer the intended word. To predict the time needed to write a word, the models proposed by Accot et al. [44] were used, which later were revisited by Cao et al. [45].

Apart from the different tasks, other properties can also be used to qualify interactions. As mentioned earlier, pointing refers to the task of placing the device in control space so as its representation in display space is positioned at a desired location. When the space of control and display are coupled or superimposed, the interaction is qualified as direct. In the opposite case, the interaction is qualified as indirect. Pointing with a mouse on a computer is thus referred to as an indirect interaction, the hand of the user and the representation of the mouse pointer are not collocated, while pointing on a mobile device is direct, the user touches the location where the target is represented. Motions performed in both interactions are similar, the main difference relies in the extra effort required from the users to mentally couple their motions to the visual feedback, a process which remain difficult [46]. Fitts' law, which was originally developed to predict movement time in direct pointing, has also been shown to be robust for indirect pointing. However, specific interactions have been shown to be more adapted to one or the other mode. Forlines et al. [47], for instance, showed that users were performing better in bimanual tasks and single-point interactions in indirect and direct fashion, respectively. Finally, interactions can also distinguished by whether they are qualified of relative or absolute. This relates to the mapping between the user pointer and its representation in display space. The relative mapping links the displacement to the representation, by opposition to the absolute mapping with establishes a correspondance. For a relative mapping, a control/display gain (CD_{gain}) of the interaction is defined, as in [48, 49]. Mouse interaction is thus prototypical of a relative mapping, where clutching is possible [50], while touchscreen interaction is representative of an absolute mapping where the position of the pointer is equivalent to its representation. Note that all four possible combinations of relative, absolute, direct and indirect interactions exists.

In this thesis, gestural interactions based on successive target acquisitions and indirect gesture typing will be proposed, and Fitts' law is employed to foresee some of the interaction properties.

Different effects can be expected when gestural interactions implement these tasks. The information throughput in pointing tasks depends on the pointing technique and/or device being used. But using different limbs has also been shown that have an impact. Card et al. [51] compiled several Fitts' study involving different body parts, (Figure 2.2 (left)), in order to produce a fair comparison. They showed that the information throughput achieved by users was higher when they used their fingers rather than their wrists or arms, observing that the performance of a device is more or less "set by the muscle groups with which the device is designed to connect". Focusing the range of required motions, effects of scale were also investigated with the hypothesis that the performance would exhibit a U-shaped curve with greater performance for medium values of scale. This was measured in steering tasks by Accot et al. [52], (Figure 2.2 (right)). Their results confirmed a U-shaped performance-scale function, although the impact of scale was less than that of the steering law's index of difficulty. The change in performance over a range of scale varying by a factor 16 was evaluated to 17%. Movement scale can recruit different limb segments: a large movement tends to be carried out primarily by the arm, while a small movement will recruit only the fingers. As a result, it has been argued that scale should also be considered as "the basic dimension of aimed movement", as pointed out by Guiard [53].

In the rest of this thesis, some of these tasks will be employed for upper limb gestural interactions. We expect thus to observe the effects described above. In particular, it is clear that an interaction that is moved from one limb to another will observe a performance change in terms of information throughput. It is also important to note that effects of scale have an impact on performance throughput as well: when limit of reach or body precision are reached,

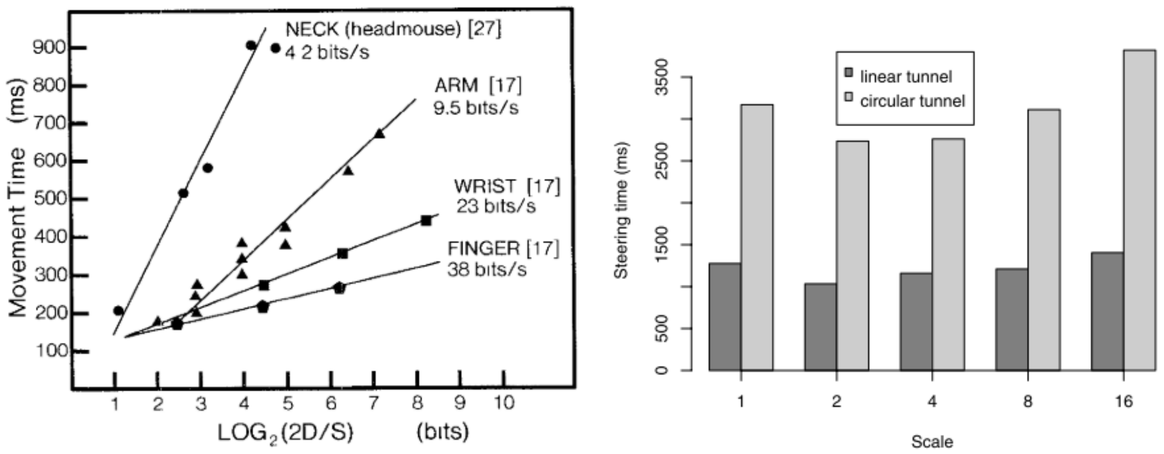

Figure 2.2: Known effects on IP in gestural interactions. Information throughput afforded by different body parts in a Fitts' law task [Card - 1991] (left). Effect of different scales for same ID on steering time in a tunnel task, exhibiting the U-shaped performance curve [Accot - 2001] (right).

the performance suffers.

2.4 Computational Interaction

The previous section has presented gestural interactions through their properties and characteristics. In particular, idiomatic tasks have been introduced such as pointing, steering or gesture typing. These were shown to obey Fitts' law in their overall majority. It is important to note that some of the experiments from which this law was derived did not involve any computing devices. In the original paper from Fitts [43], pins were transfered from one set of holes to the other or disks were moved from one pin to another in a left to right fashion. Despite the obvious differences between moving a disk and clicking an icon on a desktop computer, Fitts' law still accurately describes the observed interaction. However, because systems of increasing complexity are now heavily used, more generic models are needed in the field of HCI.

One such attempt has been proposed by Williamson [54], where the interaction of a user with a system is portrayed as a continuous control process in which user intentions are sensed through their motions via sensors, which outputs are transformed by the system for the control of the system's state variables. A simple schematics is reproduced here for illustration purposes, see Figure 2.3. One of the added benefit of using this model is that is formalises the place of the sensors within the interaction and makes explicit the transformation operated by the system between the sensors output and variables that the user tries to control. This description is heavily drawing from the field of Information Theory, the author stating that "the main purpose of the interaction is to convey information".

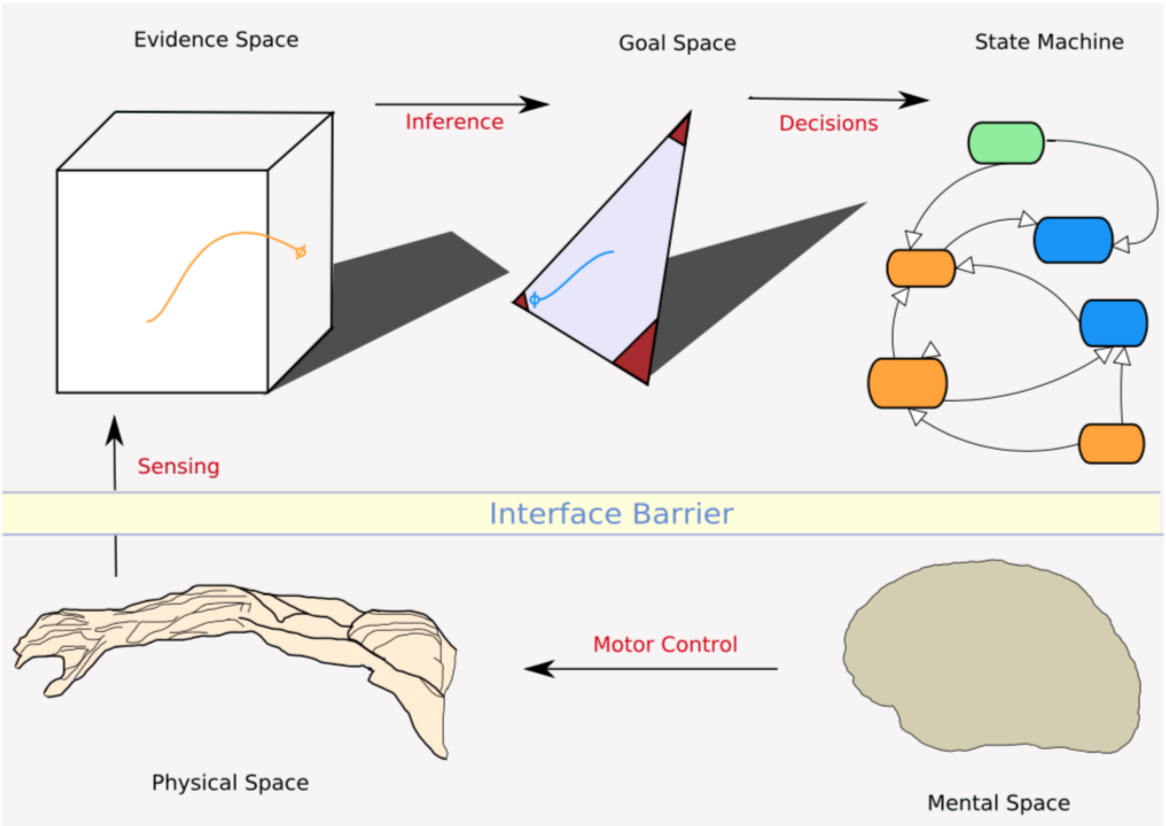

Figure 2.3: The interaction of a user with a system is decomposed into the user motions, the evidence space, the goal space and the state machine with a feedack closing the loop. Note that the feedback loop is omitted, but the human computer interaction is typically viewed as a closed-loop control process. Reproduced from [Williamson - 2006].

Several ideas flow from this description. First of all, it is well adapted to describe gestural interactions, as it explicitly includes the user's body in the model. Second, it emphasises the fact that the interaction should be regarded as a closed-loop control process and as such stresses the importance of the feedback mechanism. This idea has long been established and research has shown that humans are usually behaving so as to recreate a first-order system overall [55]. Finally, it places an important place on the uncertainty in the interactions. At every stages, inherent noise can be present. Human motions are noisy in their nature, which was the basis for establishing Fitts' law, but so is the sensing afforded by the sensors, and the processing that govern the inference mechanism.

Applied to gestural recognition such modelling pave the way for a computational approach. Many different statistical models have been employed to transform motions into a material adapted to inference. For instance, Bevilacqua et al. [56] proposed to train Hidden Markov Models [57] on movement examples to "follow" and locate where in the recorded movement a sample from an unseen repetitions was taken from. Using nonlinear dynamical systems, Ijspeert et al. [58] proposed the “dynamical movements primitives” which, after being fitted to movement examples, were capable of producing similar movement to the example as well as variations governed by different scale and goal parameters. For the purpose of computing the information content carried in a movement, Oulasvirta et al. [59] proposed to fit an autoregressive model to a example of motion and its reproduction in order to analyse the differences between the models’ coefficient which were related to a measure of information. Finally, using biomechanical simulations, Bachynskyi et al. [60, 61] proposed to analyse the muscle coactivation in the body of users engaged in different interactions. Different properties of motions are afforded by these models, such as the capacity of following a prerecorded motion or generating subtle variations of it, or assessing their quality from an ergonomic point of view. The computation of performance remains however complex, as already pointed out by MacKenzie et al. [42], “there are classes of movements (e.g., drawing) that at present lack a paradigm for performance modelling.”

While Computational Interaction [62] is still in the process of being formalised, the main idea lies in the “commitment to computational models that gain insight into the nature and processes of the interaction itself”. The proposed definition for Computational Interaction lists elements that such approach would typically include:

1. an explicit mathematical model of user-system behaviour;
2. a way of updating that model with observed data from users;
3. an algorithmic element that, using this model, can directly synthesise or adapt the design;
4. a way of automating and instrumenting the modelling and design process;
5. the ability to simulate or synthesise elements of the expected user-system behaviour.

As a result, Computational interaction draws from the fields of Machine Learning, Signal Processing, Information Theory or Control Theory. The main goal of this thesis consists in exploring how such approach can be applied to novel gestural interactions of the upper limb on planar surfaces for which the lack of models challenges their to creation, optimisation or understanding.

2.5 Conclusion

The background section has introduced the context in which this thesis has been carried out and established the connection between the findings of research on users with spinal cord injuries and the design choices of made in the rest of this thesis. Upper limb interactions have been portrayed through examples taken from the fields of surface computing and interactive tabletop, followed by an emphasis on related work based on optical tracking. General models of gestural interaction have been presented with known effects of scale and different limbs on performance. Finally, a short introduction has been provided for what is defined as Computational Interaction and the need for modelling interactions.

The following chapters of this thesis are constituted by three research projects, linked by a continuous dialogue with occupational therapists, where new upper limb gestural interactions are proposed and a computational approach is employed for the purpose of enabling, optimising or understanding the interaction.

Chapter 3

Gesture Typing Through Virtual Surfaces

Summary. This chapter proposes a novel upper limb gestural interaction whose purpose is to afford text-input through an optically tracked surface. Recent designs in the field of optical tracking have investigated a wide range of surface type (hands, walls or notepads) but report some limitations with regards to their ability to detect whether the user's finger is in contact with the optically tracked surface. A supervised learning approach based on the processing of depth images is shown to produce competitive results with the current literature, characterised by a 96% AUC (area under the ROC curve). A user study was designed to investigate the influence of the dimensions of the input surface on a gesture typing task. It demonstrates that the proposed system allows users to perform text-input, albeit at a lower rate than on a control touch tablet. Users adapted their end-effector movement speed to the input size in order to maintain similar writing speed across sizes, but reached their precision limit for small dimensions. Also, users expressed a preference for an interaction size that is big enough to limit errors but small enough to minimise arm motions. Finally, from the trace data gathered during the experiment, links between observed behaviour and control theoretic models of target pointing are established.

3.1 Introduction

An initial meeting with occupational therapists (OTs) in the spinal unit of Glasgow's Queen Elizabeth University Hospital inspired the design for the first upper limb gestural interaction this thesis proposes. A series of tools crafted by OTs and used in the ward were presented, see Figure 3.1 and in Annexe A. These were custom-built for individual patients and their purpose was mainly to alleviate the shortcomings of their hand motor limitations. The tool on Figure 3.1 was designed to allow users to hold a pen while exerting minimal squeezing forces with their fingers, whereas the tools presented on Figure A.2 and Figure A.3 were destined to facilitate interactions with touch sensitive surfaces by replacing a user finger with a protruding extension or removing the need for the continuous muscle activation which keeps a finger extended. The focus on touch interactions and writing is understandable considering how these are important in everyday life. Furthermore, text-input is also potent in the digital version of everyday life: a longitudinal study of mobile interactions pointed out that text-input apps, such as mail or messaging, accounts for as much as 40% of total phone use time [63].

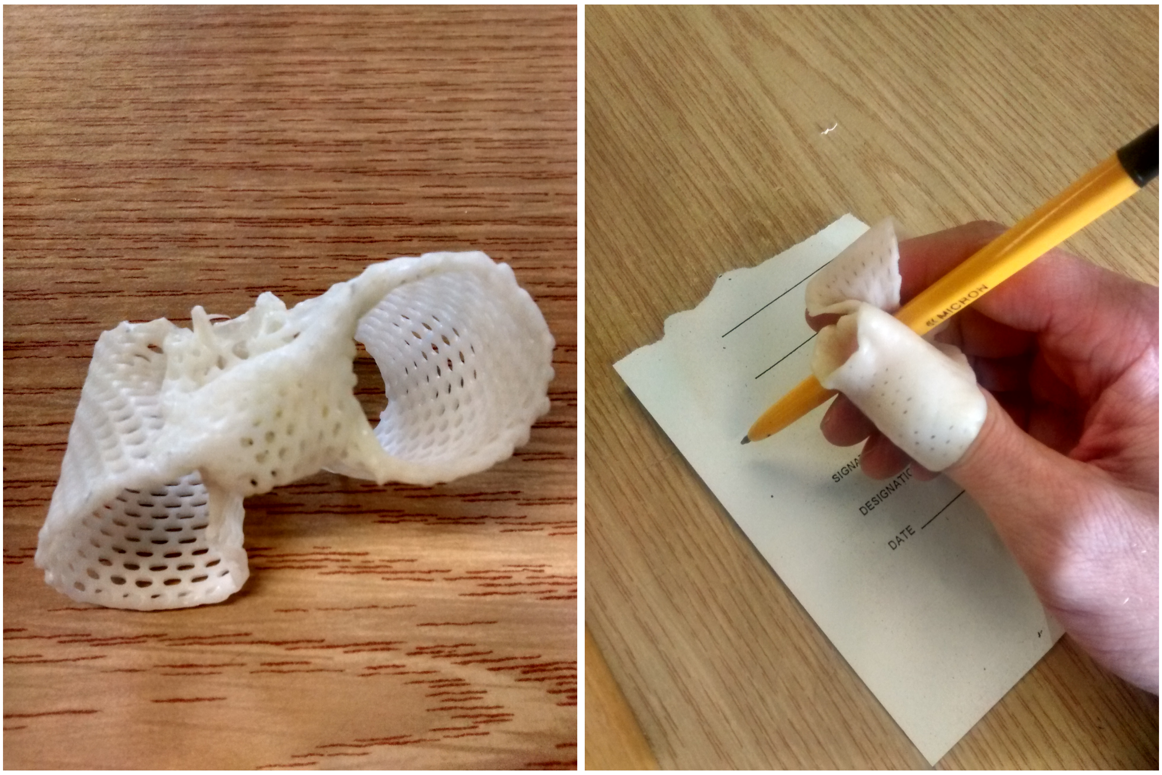

Figure 3.1: Photographies of the splint created by the OT to enable users, who could only exert a limited force in their fingers, to hold a pen or a pencil. The material used is a plastic-based deformable paste that can be shaped and set in form with warm water.

The production of these contraptions echoes the findings from the study involving users with limited mobility, see 2.1. Mobile text-input was identified as a challenge, and mobile interaction in general has been proven problematic due to the small size of screens and the likely production of spurious touches. By putting these objects in relation with the current literature on gestural interactions and optically-tracked surfaces, a potential design comes to mind that would focus on the input side of mobile interaction. An optically tracked virtual surface could extend the usable touch surface of the screen to a tabletop onto which the user hand would rest comfortably. The virtual nature of the surface could also be used to filter the potential spurious touches in the design of the signal processing and to limit the reliance on consecutive target acquisitions, gesture typing could be employed in lieu of conventional mobile keyboard interaction.

3.2 Proposed Interaction

The proposed interaction for our optically tracked surface is represented in Figure 3.2. The main artefact can be described as a marker-free, optically tracked surface that affords gesture typing. A mobile device equipped with a camera opportunistically creates an additional touch surface on the tabletop that supports it. The computing device tracks touch events on the virtual surface and uses them as an input stream, equivalent to the ones produced by its own touch screen, for driving the interaction. A user would engage in a gesture typing task by motioning their finger on the virtual surface, acting as if it was a conventional touch screen.

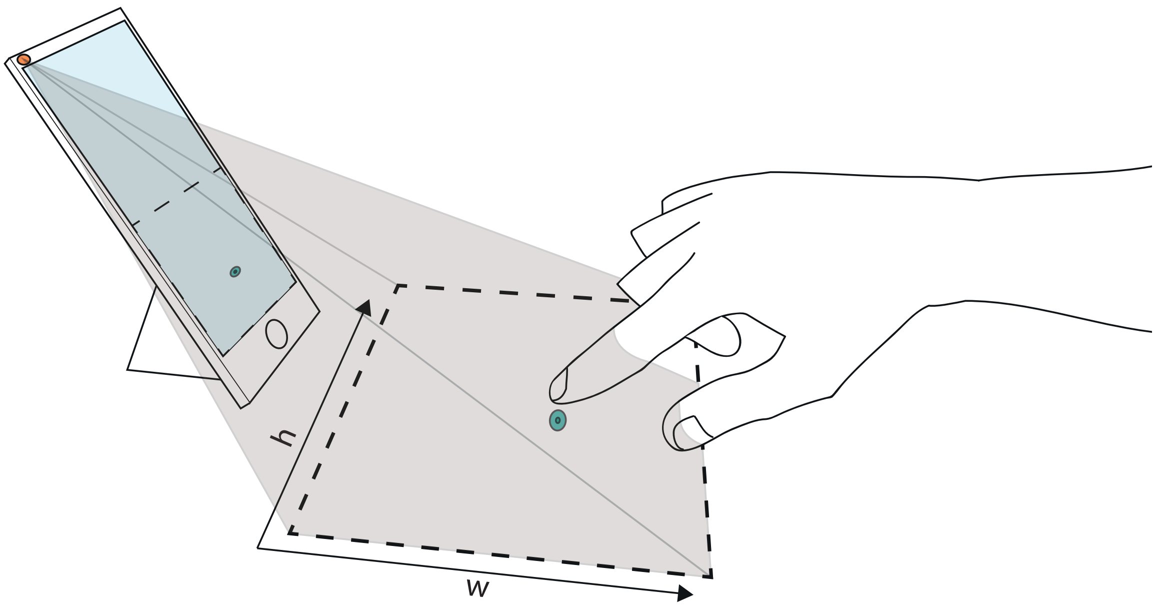

Figure 3.2: Idealised interaction model: a mobile device creates on-demand touch capable virtual surfaces using visual tracking, which enable an alternative and comparable interaction to the one afforded by its own touchscreen. The interactive surface is represented here for illustration purposes but no visual feedback is provided to the user on the tabletop.

The interactions proposed by the touch screen and the virtual surfaces exhibit very similar properties. They both offer an absolute mapping between the user pointer and the pointer in control space and rely on the touch paradigm for interaction. Both interactions also require a single touch point for interaction. However, the control/display gain (see 2.3) can be different since the virtual surface is defined by variable shapes or sizes which are fixed through two free parameters h and w representing the height and width of the input surface.

Moreover, one of the main difference between the screen's surface and the virtual surface is that the interaction they propose is direct for the former and indirect for the latter, meaning that in the virtual surface case the touch point and the feedback point are not collocated in space. Thus it is necessary to provide the user with a visual feedback of its hover and touch position at all time as indicated by the green dot on the mobile screen in Figure 3.2. This difference in directness is most likely to change the interaction from a open-loop to a closed-loop interaction.

3.3 Research Plan

Despite the apparent simplicity of the proposed interaction, several challenges can be identified.

1. Little is known about the performance of optical trackers as measured by Fitts' law tasks.
2. Continuous interaction, such as gesture typing, has not been demonstrated with such systems.
3. Tracking algorithms have been published but lack performance or availability.
4. Gesture typing has rarely been tested in an indirect manner[1], and the influence of CD_{gain} or scale is unknown.

Section 2.2 provided some background regarding related work in the context of upper limb interactions through optically tracked surfaces including a vast range of designs in that space [1, 31, 32, 33].

Experiments that have been carried out in the current literature have focused on estimating the system's tracking performance in terms of the positional error and touch classification. These are undertaken in a open-loop fashion where movement times are not measured or reported. As explained in section 2.3, classical system evaluation usually measures the movement time as part of a Fitts' law modelling, which then provides a measure for the available information throughput. This matters for evaluating the performance of a given input modality and comparing it to other input techniques. It is also important to record the interplay between the users and the system, as users become engaged in a trade-off between speed and accuracy, balancing the act of doing the task correctly and as quickly as possible. If the participant is given an unlimited amount of time, there is no bound to the accuracy that can be produced and this could give a false sentiment for how well someone would perform.

[1] This has recently been investigated more thoroughly by Yang et al. [64].

The performance of these systems remains far from those of conventional touch screens. For example, the classification of touch for the user pointer yields a rate of 3.5% false negative and 19% false positive in [33] while the positional errors in tracking are reported in the range of tens of millimetres in [32], one order of magnitude higher than the human performance as measured by Holz et al. [65] in the range of the millimetre. Several reasons explain these differences. First, as opposed to capacitive touch screens, the sensing in optically tracked surfaces usually happens at a distance, ranging from several centimetres up to several meters in extreme cases. As the distance between the user's pointer and the sensor increases, the number of pixels representing the user's pointer decreases, reducing the amount of sensed information that can be used to infer its position and potential contact with the surface. Second, the techniques employed for the touch classification have used heuristic methods for solving a complex problem where algorithmic parameters could be instead inferred from the data itself by using techniques from the field of machine learning. For instance, the technique presented by Harrison and Xiao relies on the thresholding of the height value of a patch of pixels surrounding the intended user pointer, with the threshold arbitrarily fixed. Thirdly, as depth cameras remain a relatively recent technology, research has relied on early prototype versions of hardware far from customer-ready devices. As a result, such systems are still at best prototypes for research in laboratories.

Currently, there exists no available tracker that would allow for the investigation of these questions. Popular pose tracking system (Leap and Kinect for example) do not fare well with a tabletop in their field of view. Furthermore, the need for design of a model of touch classification also limits the possibility of using systems which do not provide access to the raw camera data. Finally, in the optics of a marker-free design, some systems such as Optitrack can not be considered. Therefore, we set out to build a simple prototype.

Beyond tasks principally aiming at measuring the performance of a tracking system, other ecologically valid tasks can be used. As mentioned in section 2.3, aside from pointing and steering tasks, gesture-based tasks can also be the basis for valuable interaction techniques. We propose that gesture tasks constitute good candidates for optically tracked surfaces, in particular gesture typing. Gesture typing is a relatively novel text-input technique that was originally introduced by Kristensson et al. [5]. In order to write a word using this technique, a user motions their pointer without interruption over the word's constituent letters. Thus, this design would mitigate inherent system errors, as gestures also encode information in the relationship between samples and the decoding can usually cope with a high degree of positional uncertainty on individual samples. Moreover, in opposition to conventional text-input, the reliance on the touch classification is greatly limited as most of the interaction time is spent in hovering or touching state, but transition from one state to the other only happen during the target acquisition of the first letter. There exists some closely related research on gesture typing in indirect interaction or on the influence of scale on gesture typing, each providing some prior pieces of information that will be used in the analysis of results. This includes *Vulture*, the system Markussen et al. [66] proposed, that afforded gesture typing in an indirect manner by having users motioning their finger in a mid-air interaction. Vertanen et al. [67] have investigated the influence of input dimension on a gesture typing task, aiming at understanding how small could the interactive surface be and whether a smartwatch could be used as input.

In this chapter, we are interested in exploring the use of a novel upper limb gestural interaction on a planar surface. From the literature, we gathered an interaction potential for optically-tracked surfaces albeit with some challenges in term of tracking and touch classification.

This chapter aims to answer the following research questions:

- **RQ0.1**: How can computational models be used to transform image data into a material enabling touch interactions?
- **RQ0.2**: How does this new interaction compare to the control condition of a tablet with capacitive touchscreen?
- **RQ0.3**: What is the influence of size on the text-input performance and on the interaction perceived quality?

3.4 Implementation

The physical setup for our interaction was composed by a depth camera mounted on a tripod that overlooked a desk table onto which a mobile device was resting vertically (Figure 3.3). The virtual surface is created just in front of the mobile device. We used an Intel Realsense SR300 depth camera for this purpose. On the software side, the processing steps involved for creating a virtual surface start with the definition of a coordinate system with respect to the sensed environment. This is followed by the detection and tracking of the user's pointer position within the coordinate system, and the modelling of its categorical touching property. In other words, the goal of the processing pipeline is to transform an input stream of colour and depth images into a vector (x, y, t) containing the position coordinates (x, y) of the user's pointer over the surface and a variable t indicating whether the pointer is in contact with the surface.

3.4.1 Processing Pipeline

The first step is derived from techniques widely used in augmented reality [68]. A random pattern was used to indicate the desired position of the virtual surface on the tabletop and es-

Figure 3.3: Layout of the prototype in-situ. A mounted camera overlooked the interaction area created in front of the mobile device.

tablish a frame of reference. The homography between the pattern and the view in the colour aligned depth stream is computed. This provides the perspective transformation between the marker and the captured image, which represents the orientation of the camera with respect to the scene. Despite the computation not indicating the scale under which the view was seen, the additional information provided by the depth image allows a unique solution to be found for the camera orientation. In other word, a single RGB-D image of the pattern on the tabletop permit to compute the absolute position of the virtual surface and attach a Cartesian coordinate system $(0, \vec{x}, \vec{y}, \vec{z})$ to the virtual surface. This technique relies on the presence of a pattern in the field of view during the setup phase, which could be alleviated with different techniques such as those proposed in [31], but this is not the focus of the presented work. After computation of the reference frame, the marker was removed and the following interactions were considered as marker-free. Figure 3.4 shows a 3-D representation of the depth data at this stage. In purple are represented all unclassified points, and in blue the points that fall within the boundaries of the virtual surface. The coordinate system indicates the surface's origin and orientation.

The following steps involved the tracking and classification of the user's pointer. As mentioned in 3.2, due to the nature of the proposed interaction (gesture typing), we are interested in single-touch interaction only. As a result and for the sake of simplicity, we avoided modelling the hand pose or identifying individual finger to infer the intended user's pointer[2]. Instead, we adopted the convention that the closest protruding pixel detected as hovering over the surface plane would belong to the user's intended pointer and be used as the seed for the further processing. As expressed within our coordinate system, a region of interest whose limits are defined by the vertical cylinder of five centimetres in diameter that surrounds the seeding pixel is selected. Figure 3.4 depicts the region of interest in cyan and all the points within a centimetre of the seeding pixel in red. The pixels that are contained within the cylinder but extend further away than one centimetre from the seeding pixel are marked in orange.

In practise, this design choice forced the user to interact in a front-facing manner with the system. Indeed, when the hand is facing the camera, our simple algorithm ensures that the user pointer was correctly detected. Preliminary informal testing has shown that this constraint did not prove to be an issue as the motions required by gesture typing did not induce tilting of the wrist beyond a point where the palm of the hand, for example, would find itself as the closest protruding point to the camera. For subsequent tests, this paradigm was made explicit and clearly stated beforehand to the participants. Finally, even if this design choice represents a constraint, it also provides some opportunities by allowing users to interact with any number of bundled fingers or with a hand posture they find comfortable.

This processing step produced a segmentation of the input stream in different pointclouds representing the virtual surface boundaries, the volume of interest within which the interaction will occur and inferred intended user pointer, (Figure 3.4). The remaining information that was extracted was the pointer position and whether the pointer was touching the virtual surface. These two pieces of information were essentially a regression and a classification performed on the pointcloud representing the user pointer and are the subject of the following sections.

3.4.2 Touch Regression Model

The regression model is aiming at producing the (x, y) coordinate of the user pointer with regards to the coordinate system and takes as input the pointcloud representing the user pointer produced by the preceding step. The mean value of the pointcloud is first taken as an estimate for the fingertip position. This resulting point could exhibit two types of errors in the form of an offset associated with a noise. With regards to the offset, since the intended interaction is performed in closed-loop, users will continuously adapt to discrepancies between

[2] Hand pose modelling is a very active area of research and the latest advances such as openpose [69] which software package is now available on https://github.com/CMU-Perceptual-Computing-Lab/openpose have demonstrated outstanding and accessible results even though its expensive computing cost currently prevents applications in mobile real-time interactions.

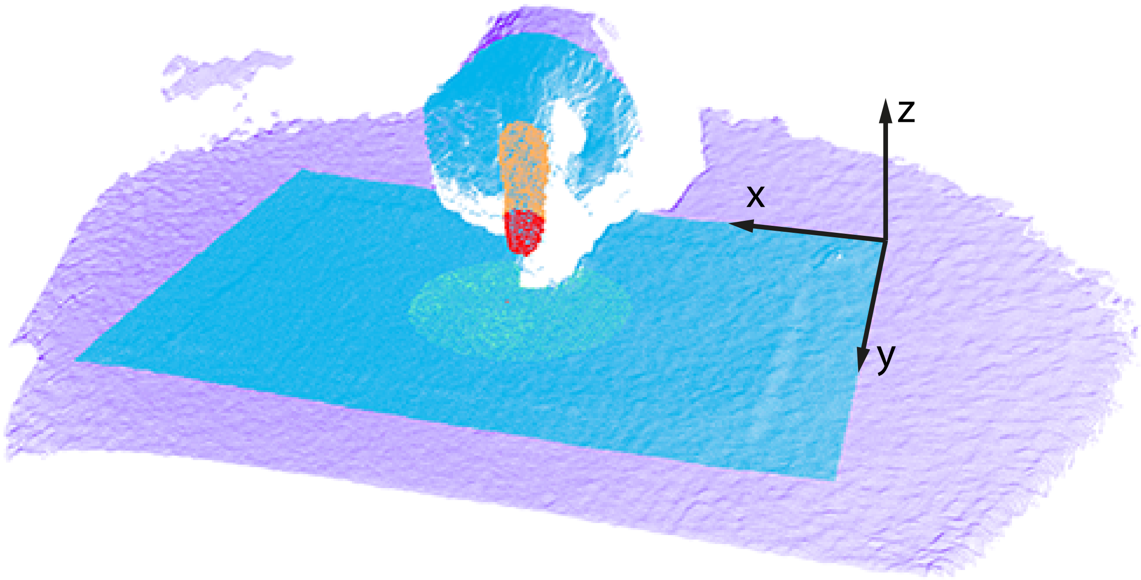

Figure 3.4: Schematics of the output of the processing pipeline with colour-coded segmentation output.

the visual feedback they receive and their intended position. As a result and provided the offset is constant, only one correction at the beginning of the interaction is needed to cancel it. Regarding the sensor noise however, its unpredictable nature calls for a more complex model.

A Kalman filter [57] was chosen to model the evolution of the estimated position in time. It is the optimal estimator assuming the noise is Gaussian. The Kalman filter is an algorithm that uses a linear dynamical system $\mathbf{F}_k$ to describe the relationship between state variables $\mathbf{x}_k$ and observed variable $\mathbf{z}_k$ at time k for regular intervals in time. The state variables are what we model, while the observed variables are what we measure. This algorithm permits to represent noise as an integral part of the system and to track the uncertainty of the measurement.

In mathematical terms, the following relation hold true for the hidden state $\mathbf{x}_k$, its precedent estimate $\mathbf{x}_{k-1}$ and the process noise $\mathbf{w}_k$:

$$\mathbf{x}_k = \mathbf{F}_k \mathbf{x}_{k-1} + \mathbf{w}_k$$

where:

$$\mathbf{w}_k \sim \mathcal{N}\left(0, \mathbf{Q}_k\right)$$

follows a normal distribution with zero mean and covariance $\mathbf{Q}_k$. The relation between the hidden state and the observed state is governed by the observation matrix $\mathbf{H}_k$ and the observation noise $\mathbf{v}_k$:

$$\mathbf{z}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k$$

where:

$$\mathbf{v}_k \sim \mathcal{N}(0, \mathbf{R}_k)$$

is assumed to follow a normal distribution with zero mean and covariance $\mathbf{R}_k$.

For our purpose, the hidden state includes the position and the velocity of the pointer as parameters, which lead to the constant state-transition model as:

$$\mathbf{F}_k = \begin{bmatrix} 1 & 0 & dt & 0 \\ 0 & 1 & 0 & dt \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Including the second derivative would be possible, however, the sufficient quality of the results in term of tracking did not seem to warrant more complexity. The observation matrix, which transform the hidden state into the observed variable is:

$$\mathbf{H}_k = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

This filtering step is meant to take into account the unavoidable human errors in tracking as well as the tracking errors in the processing chain, quite common with optical system. These two sources of noise are thus associated with the process noise and the observation noise.

Experiment

The free parameters are the process noise and observation noise through their respective covariance $\mathbf{Q}_k$ and $\mathbf{R}_k$ which reflect the noise present in the measure of the observed variable and the noise present in the hidden state, respectively. Because the values for these parameters are difficult to infer from the data, these are traditionally specified by hand. Here, we decided to fix the transition noise to unity and vary the observation noise as the identity matrix multiplied by a increasing scaling factor, assuming that the human noise in positioning is lower than the tracking noise.

To find a suitable value for the observation noise, a small experiment was designed with one dependent variable (the scaling factor of the observation noise) taking three different levels (0.0, 1.0 and 7.0). We recorded data from three participants engaged in a shape tracing task similar to the one used in [32], albeit in an indirect fashion. Participant were asked to follow a shape on the screen. A circle of radius $150px$ on the screen was used, corresponding to $5.21cm$ on the tabletop. We instructed participants to follow the circle outline for one minute while being as fast and accurate as possible.

The shortest distance to the circle periphery was computed for each point and a normal distribution was fitted on the resulting distribution. The results are presented in Table 3.1, with associated distribution depicted in Figure 3.5. The overall average positional error was less than one millimetre for all conditions. The mean value of the error appeared to decrease with an increase of the level of observation noise. A mean value close to $0.0mm$ was expected as participants continuously corrected their position with positive and negative errors cancelling out on average. In terms of performance, the width of the distribution is potentially more interesting than its mean as it indicates the ability of participants to stay close to the target. For the level with no observation noise, a standard deviation of $2.0mm$ was calculated, while a standard deviation of $1.5mm$ was observed when the observation noise was set to 1.0 or 7.0. Because an increase in observational noise presents a drawback: it increases the lag in the system, we chose to set the threshold for the observation noise to 1.0.

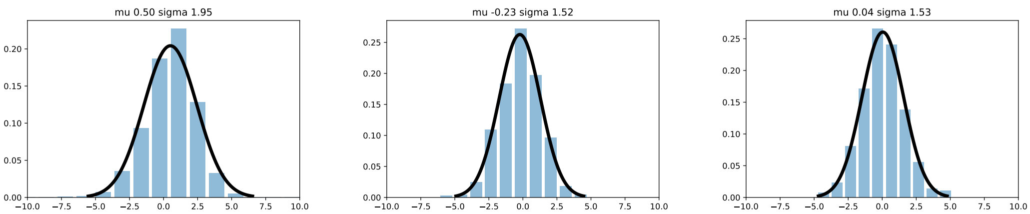

Figure 3.5: Effect of Kalman filtering and observation noise on the tracking precision in a shape tracing task.

$\mathbf{R}_k$	0.0	1.0	7.0
$\mu[mm]$	0.5	-0.23	0.04
$\sigma[mm]$	1.95	1.52	1.53

Table 3.1: Mean value and standard deviation for the position error in millimetres for different values of the observation noise $\mathbf{R}_k$ in the Kalman filter.

It should also be noted that this tracking performance applies only to the display feedback, but does not insure that the tracking is accurate in the control space. The camera we used produced a slightly distorted depth image causing a planar surface (an office desk) to be represented as slanted in the lower left corner of the image. Optical distortion correction such as those proposed in [70] were not applied since the observed distortions were very localised and judged relatively small in comparison with the interaction scale. Moreover, the closed-loop nature of the interaction allowed users to correct for such inconsistencies. As a result, the user positional error we measured from the regression model were smaller than those previously reported in [33] in which the mean value and standard deviation in position error were $4.0mm$ and $3.4mm$, respectively, for the same task in an open-loop interaction.

The following section will look into the touch classification model.

3.4.3 Touch Classification Model

When using depth cameras, the main method reported in the literature [31, 32, 33] for solving the touch classification problem is to use thresholding on a patch of pixels that surrounds the detected fingertips and the latest results present a performance that remains problematic for complex tasks were continuous and sustained tracking is required. Data-driven methods, such as machine learning techniques, have recently demonstrated their suitability for image classification. The touch classification problem, which reduces to finding out whether the user pointer is in contact with a surface from an image produced by a depth camera, seems a good candidate for this task.

Supervised Learning Approach

The touch classification problem can be framed as a supervised learning approach in which we aim to find a function f parametrised by θ that relates an input vector X into a categorical output Y such that:

$$Y = f(X, \theta)$$

Framing the problem in a supervised framework is conditioned on the existence of a dataset of pairs (X, Y). The correspondence between X and Y are examples of the relationship we wish to model. This framing allows an optimisation to take place where the cost function $L_\theta(X, Y)$ depending on the θ parameter is to be minimised. In other words, we are looking for the value $\hat{\theta}$ of parameter θ that solves:

$$\hat{\theta} = argmin_\theta(L_\theta(X, Y))$$

There are different models for the mapping function f and the choice of model mostly depends on the nature of X, whether the input data are images, time series or else, the supposed relationship between X and Y and whether we wish the model to be interpretable or not. In the case of touch classification, the input vector X is an image recorded by a depth-camera containing a possible touch point while the output Y indicates whether a touch point should be detected in this image. Models based on neural networks [57] have recently been very popular and successful solving this kind of problem. Depending on their complexity, neural network can require a lot of data for their training, typically in the order of thousands of images. The following section will thus delve into the properties of our dataset after detailing how it was collected.

Dataset

Currently, there is no available dataset on which to rely to train the classification model, refer to [71] for a non-exhaustive list of RGB-D datasets. A suitable dataset would qualify if it had been produced with the same or comparable optical sensor, presented similar or comparable viewing angles and included the labels we need to model. Even data from related research studies [32] would not qualify as the sensor used were different from the one used for this research. As a result, we propose to capture and create a dataset to be used in the training of the classification model.

The collection and labelling of data can be made practical. To avoid having to segment videos and annotate individual frames into our two classes; touching and non-touching, both categories were recorded separately. To record touching category data, a continuous video stream of a finger in contact with the surface was captured. The non-touching category was captured as a continuous video stream of a finger hovering at different height over the surface (from less than a millimetre to several centimetres). The labelling was performed on a video stream basis, which allowed the annotation of hundreds of images at a time. The quality of a dataset depends on its capability to represent all the cases that the model will need to classify. As such, in the touching category we ensured that different orientation of the front facing hand were recorded and the non-touching category included frames in which the hand was out of view. For the same reason, since the dataset was produced by the same individual, the interaction with different fingers (thumb, pinky and index) were recorded to emulate different potential users. Finally, the sensor placement was established at two different distances from the surface (at roughly $40cm$ and $80cm$ from the surface centre) to account for varying camera poses. A special attention was placed on the filtering software afforded by the camera. As it tends to remove some valuable artefacts containing information about touch, it was turned to a minimum.

A total 6 minutes of video data was recorded, evenly balanced between touching and non-touching categories, camera views and fingers. The final dataset contained 2 categories $\times$ 2 camera views $\times$ 3 fingers $\times$ 30 seconds $\times$ 30 frames per seconds, equivalent to 10800 datapoints.

Features

From this dataset and to simplify the task of the model f, some features were extracted. The output of the processing pipeline from 3.4.1 was used as a starting point for the feature engineering. In the detected user pointer pointcloud, points which were lying between $-1cm$ and $3cm$ along the $\vec{z}$ axis were selected. Their distribution was computed and discretised into 20 uniform bins. The distributions of the two categories for one video recording is shown in Figure 3.6. The difference between the distributions is quite apparent with a density for the touching category between feature dimension $\#5$ and $\#15$ much higher than non-touching category. Note that feature $\#5$ corresponds in the $\vec{z}$ to the origin axis, where the virtual surface plane is located. The histogram operation provides a rotation and shape invariance as well as a reduction of the number of dimensions. The simplicity of the features and their invariance to rotation and shape is key to limit the amount of data that is needed for the successful training of the model.

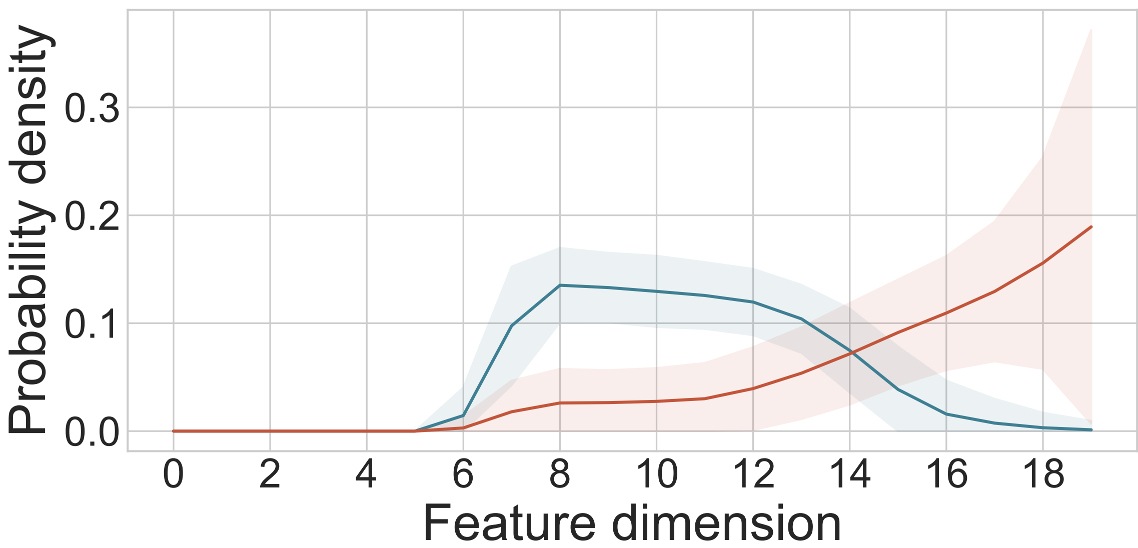

Figure 3.6: Distribution of the features for touching and non-touching frames in blue and red, respectively. The shaded area indicates one standard deviation from the mean value represented as a solid line.

Neural Network and Results

Given the features have a rather low dimensionality, only twenty, whereas images usually involves hundreds, and since categories appear to be easily separable when plotted aside each other (Figure 3.6), we opted for a neural network with a shallow architecture and without convolutional layers. The architecture is taken from code samples available with the library

Keras [72] used for the computation. The layers were the following: an input was connected to the extracted features and was followed by two stacks of a fully connected layer with rectified linear (ReLU) activation associated with a 50% dropout layer. The network is producing the output through a final fully connected layer with sigmoid activation. Refer to Table 3.2 for a full description. The architecture of neural network could be optimised for different objectives, but the satisfying performance did not warrant such operation.

Layer	Output Shape	Setting	Param #
input	20	n.a.	0
dense	64	relu	1344
dropout	64	50%	0
dense	64	relu	4160
dropout	64	50%	0
dense	1	sigmoid	65

Table 3.2: Neural network architecture with a total of 5569 parameters.

We trained our model with the optimiser rms prop [73] using the loss function binary cross entropy. We used cross validation across different fingers to verify during training the generalisation power of our model against new users. We obtained an averaged 0.96 AUC for the ROC with an operating point of 0.5, (Figure 3.7). For the live system, we trained on the whole dataset for 75 epochs.

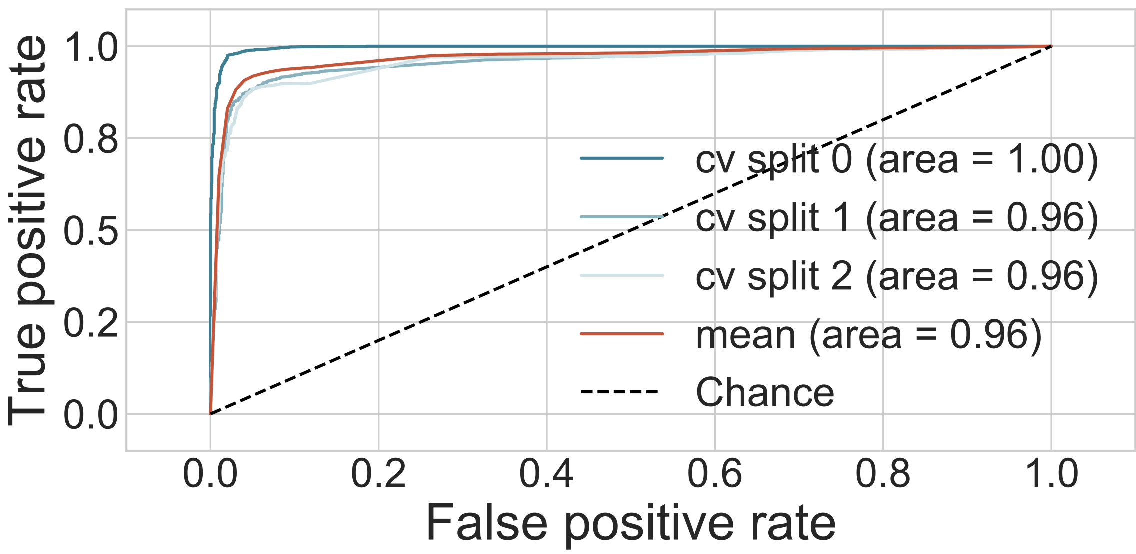

Figure 3.7: Performance of the touch classifier with 3-fold cross validation and averaging.

Finally, this model was included in a 3-states button following a design proposed by Buxton[74], (Figure 3.8), where it played the role of deciding for the transition “hand close” and “hand away” as well as “pointer lift” and “pointer down”. To improve the performance of the 3-state button and reduce further the probability of spurious touches generated by our classifier, we chose two different values for the operating point of the classifier which was dependent on the state transition. The operating point was set to 0.3 while “touching”, but 0.7 while “hovering”, effectively making the transition from touch (“state 2”) to hover (“state 1”) more difficult than the opposite.

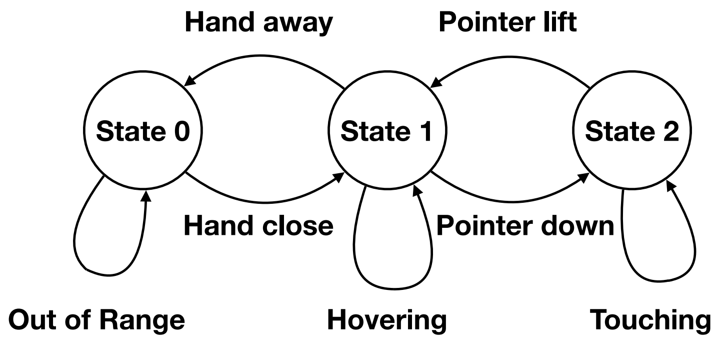

Figure 3.8: 3-state model for the virtual button.

Discussion

A fair comparison of the performance of our model with the performance of related systems such as [32, 33] is difficult to make without having access to the related systems or accurately re-implementing the proposed algorithms[3]. The differences in terms of sensors employed (Kinect versus Realsense) and associated resolution, frame rate and the differences in the sensors' placement relative to the touch point in term of distances and angles are too great to place the published results of different approaches on an equal footing. However, the quality of the results obtained here showed that a machine learning approach to touch classification is adapted. The same approach could yield improvements for relatively similar systems, such as those mentioned in the literature, which echoes what is presented as future work in [33].

3.4.4 Conclusion

This section serves as a partial answer to **RQ0** which was interested in investigating how computational models can transform image data into a material that enables the proposed interaction. The regression model with Kalman filtering provided a low positional tracking error with a mean value in the order of the millimetre. A neural network applied to specifically engineered features delivered a touch classification with $0.96 AUC$. The processing cost in term of latency is below the frame rate of the camera which runs at 30 frames-per-seconds and allows for real-time interaction. There exists a number of different models (both in regression and classification) that could be employed to turn image data into touch data, as

[3] The source code for [32] has been made available after the completion of this work, see `https://github.com/nneonneo/direct-handtracking`

well as a number of possible model combinations, and it is difficulty to provide a fair comparison with other systems in this space, due to the diversity in terms of sensing conditions. Therefore, a definitive answer is beyond the scope of this work. Nevertheless, the current implementation can serve as an experimental probe to study human behaviour in single-touch interactions afforded by marker-less virtual surfaces.

3.5 Experiment

The system developed in the previous sections affords touch-like functionality through virtual surfaces and can as such provide text-input when connected to a mobile device that supports gesture typing. The introduction has highlighted that gestural interactions, when performed with the upper limb on a planar surface, are likely to recruit a wide range of different muscles groups for the production of required motions. Using this prototype, I was able to carry an experiment to answer the remaining two research questions, re-written here for clarity:

- **RQ1**: How does upper limb interaction through virtual surfaces compare to the control condition of a tablet with touchscreen?
- **RQ2**: What is the influence of size on the text-input performance and on the perceived interaction quality?

3.5.1 Apparatus

A desktop computer equipped with a processor Intel Core i7-4790 used the processing pipeline described above and ran at 30 frames-per-seconds. The system implemented the 3-state button model [74]. For graphical feedback, a tablet of type Android and model Nexus 7 was used. Touch events were sent to the tablet running a custom application which overruled its input event system. In addition, hover state was displayed as a red marker at the pointer's position, touching was displayed through the Android debugging facility as a cross spanning the field and continuous touch was displayed as a trace. A audio feedback is produced on each touch down event. Both machines were connected via a USB cable forming a local area network with sub-millisecond latency. The tablet ran a second experiment application that displayed the target word and a standard Swype keyboard for input. Note that the user input space was only mapped to the portion of the screen that contained the keyboard (Figure 3.2). The presence of an Android tablet allowed for the use of commercial-grade gesture typing with the low-latency emulated cursor from the camera-tracked finger position.

3.5.2 Task

We used a gesture typing task. Participants were asked to input the target word presented on the tablet's screen using the gesture typing technique while being as fast and accurate as possible. The input stimuli was the target word (WORD) and was a random sample of twenty words among the fifty most common English words between two and five letters as consistent with other approaches found in the literature. The presentation of WORD was two-fold. We randomly picked five words to be always presented first and in order, as to serve as training across all conditions and then randomly presented the remaining fifteen words for each condition. The sample used for the experiment was (BADLY, SEEM, END, ASSET, CHEW) for training words in presentation order and (ASSET, BADLY, CHEW, DECK, DOG, END, FIX, GATE, HERB, HIDE, HOT, IRON, LAB, LAY, ORDER, SAFE, SEEK, SEEM, WANT, WRONG) for the testing words. This ensured the learning was consistent across participants and LEVEL.

The first choice output from the the tablet's decoder was retrieved and compared to the target word. In case of an exact match, the task was marked as successful. A picture of the keyboard used for the experiment is shown on Figure 3.9

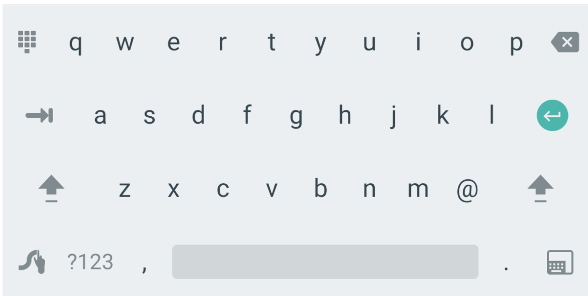

Figure 3.9: Picture of the Android's software keyboard that was presented to participants during the experiment for an orientation in portrait mode.

3.5.3 Design

Since gesture typing is not a technique predominantly used by a vast majority of users, we decided against placing some inclusion criteria related to the participant level of expertise. As such, we paid special attention to the design of the experiment block to mitigate against participants with little experience which presented the risk of confounding the experiment with a strong learning effect. We designed the block in the similar fashion to Quinn et al. [75], where the same words were entered in succession to minimize potential learning effects. It also emulated experienced behaviours by limiting the reliance of participants on gesture recall between tasks. Each trial began with a red flash of the screen and a display of the target word. A counter displayed next to the word was incremented each time a successful attempt was recorded providing a indication of performance to the participant. The block ended either after five successful inputs or seven total attempts.

A repeated measures within-subjects design was used in which three conditions and five level combinations were evaluated. The first condition was DEVICE and had two levels (TABLET and OPTICAL). This provided a fair comparison to answer **RQ0.1**. The second condition was SIZE and was evaluated with three levels (SIZE 1, 2 and 4). Size is the scaling factor in the control space and is equivalent to the inverse of the CD_{gain}. As a result, SIZE 4 presented an interaction size four times that of SIZE 1 in both dimension. The last condition was ORIENTATION and had two levels (PORTRAIT and LANDSCAPE). We adopt here a 3-symbols naming convention: the first letter represents the DEVICE, the second marks the ORIENTATION and the number represents the scaling factor as SIZE, see Table 3.3 for details. Note that for OL2, the surface area was matching that of OP2. The presentation order of the five combinations was randomised and presented a uniform distribution across the participants.

The experimental design was thus: 12 participants $\times$ 5 LEVEL $\times$ 20 WORD = 1200 trials. For each trial, we had a block of 5 to 7 ATTEMPT depending on the error rate, which equates to a total of 6,000 to 8,400 total task samples.

For each attempt in the experiment, we recorded the time the target was presented and the time the trial ended with its success condition. For the OPTICAL condition we recorded the data points sent to the tablet as sampled at the camera's frame rate.

LEVEL	*DEVICE*	*width*	*height*	*area*	*ratio*	CD_{gain}
OP1	OPTICAL	9.4	4.7	44.2	2	1
OP2	OPTICAL	18.8	9.4	176.7	2	1/2
OL2	OPTICAL	25.6	6.9	176.7	3.7	1/1.7
OP4	OPTICAL	37.7	18.9	712.5	2	1/4
TP1	TABLET	9.4	4.7	44.2	2	1

Table 3.3: Dimensions in centimetres, area in squared centimetres, ratio and control/display gain for all five combinations used in the experiment.

3.5.4 Procedure

A short introduction to gesture typing was given to the participants. Participants were asked about their previous experiences, if any, with gesture typing systems. The participants then interacted with the tracking system through a drawing application and were asked to gesture type "hello world" three times, so that they could familiarise themselves with the system. The CD_{gain} was fixed to $1/2$, the orientation was set to PORTRAIT and the mapping to the full extent of the screen, which combination was not present in the subsequent trial.

Participants were instructed to be as quick and accurate as possible when undertaking in the task. After each level, participants were offered to take a break before moving to the next one. Finally, participants were asked for their feedback using the NASA Task Load Index [76] to assess the perceived workload after each completed level. They were rewarded with a choice of candy bars on completion of the experiment. The experiment lasted on average for one hour for each participant.

3.5.5 Participants

Twelve unpaid volunteers served in the experiment: mean age of 29 (SD=5), seven males, all right-handed. All participants provided informed consent, and the experiment was approved by the University Ethics Board.

3.5.6 Results

The dependent variables were the success rate, time taken per task and trace data where available. This allowed us to compute the dependent measure error rate (ERROR RATE) defined as the percentage of unsuccessful attempts along with the text entry rate (INPUT RATE or I_r) measured as words per minutes (wpm) for successful tasks. This was computed, according to Markussen et al. [66], through the formula:

$$I_r[wpm] = |T|/s \times 60/5$$

where $|T|$ is the length of the transcribed string and s is time taken per task in seconds.

The design of the experiment indicate between 6,000 to 8,400 trials depending on participants task successes, we recorded 6963.

One of the testing word (LAY) was an outlier for the ERROR RATE across all conditions. Its mean value was at 89.5% while the WORD mean was 20.1% and no other word had an average error rate higher than 30%. A couple of test trials with the system showed that the tablet recogniser consistently promoted words of higher prior probability in the language model, such as "Larry", "last" or "Katy" over the word "lay". As a result, for the subsequent analysis, LAY was removed from the dataset.

Learning Effects

Participants were ask to report their experience with gesture typing. Half of the participants (N=6) reported having used gesture typing only once before the experiment. Two participants declared having never used gesture typing before. The rest of the participants (N=4) declared using gesture typing at a frequency between "sometimes" up to "weekly". No participants declared using gesture typing more often than weekly. This level of experience in our participants was not unexpected as gesture typing, despite being available on the most popular mobile platforms, is not always turned on by default. Some participants reported not being aware of the existence of the technique on their device. See Table 3.4 for a table summary of the results.

	never	*once*	*sometimes*	*weekly*	*more often*
# of participants	2	6	3	1	0

Table 3.4: Participants experience with gesture typing.

For the analysis of the potential learning effects, the OPTICAL condition was excluded as we were interested in learning effects with our system. We also excluded the level OP1 for the ERROR RATE as it was an outlier. The INPUT RATE and ERROR RATE were compared as a function of the presentation in the experiment of LEVEL, WORD and ATTEMPT. The results are shown on Figure 3.10.

A statistical analysis did not show an effect of LEVEL on INPUT RATE and ERROR RATE. The INPUT RATE did present an average mean value of $13.6wpm$ with a standard deviation of $3.1wpm$ while the ERROR RATE did have an average mean value of 10.8% with a standard deviation of 7.4%. This result demonstrates that the design of the block with consecutive repetitions of the same word, as used by Quinn et al. [75], did mitigate the learning effect across the experiment even when participants were inexperienced.

A more granular analysis of the experiment showed that WORD produced different performance in terms of INPUT RATE and ERROR RATE, in the same manner as it was made very apparent from the result with the WORD outlier LAY. We observed a strong difference between the first five training words and the next fifteen testing words. For training words, the mean INPUT RATE averaged at $17.6wpm$ with a standard deviation of $4.4wpm$ whereas for testing words the mean averaged at $16.6wpm$ with a smaller standard deviation at $0.4wpm$. In other words, fluctuations across the training words were observed, after which the random presentation of the testing words reduced drastically the variance. Similarly, ERROR RATE exhibited more variance on the training words than on the testing words with a standard deviation at 5.0% and 2.0%, respectively. In addition, there was a decreasing trend in ERROR RATE throughout the training words, with the first WORD having on average 26.0%, far from the mean value of the training words which was 14.6%. For the rest of the analysis, the training words were excluded.

Finally, the influence of the repetition in the experimental blocks was investigated. A statistical analysis showed a significant effect of ATTEMPT on INPUT RATE ($F_{2.21,24.35} = 32.75$, $ges = 0.14$, $p < 0.0001$). Pairwise comparison showed that the first attempt was significantly lower than the rest of the attempts with $10.8wpm$ when compared to the average of

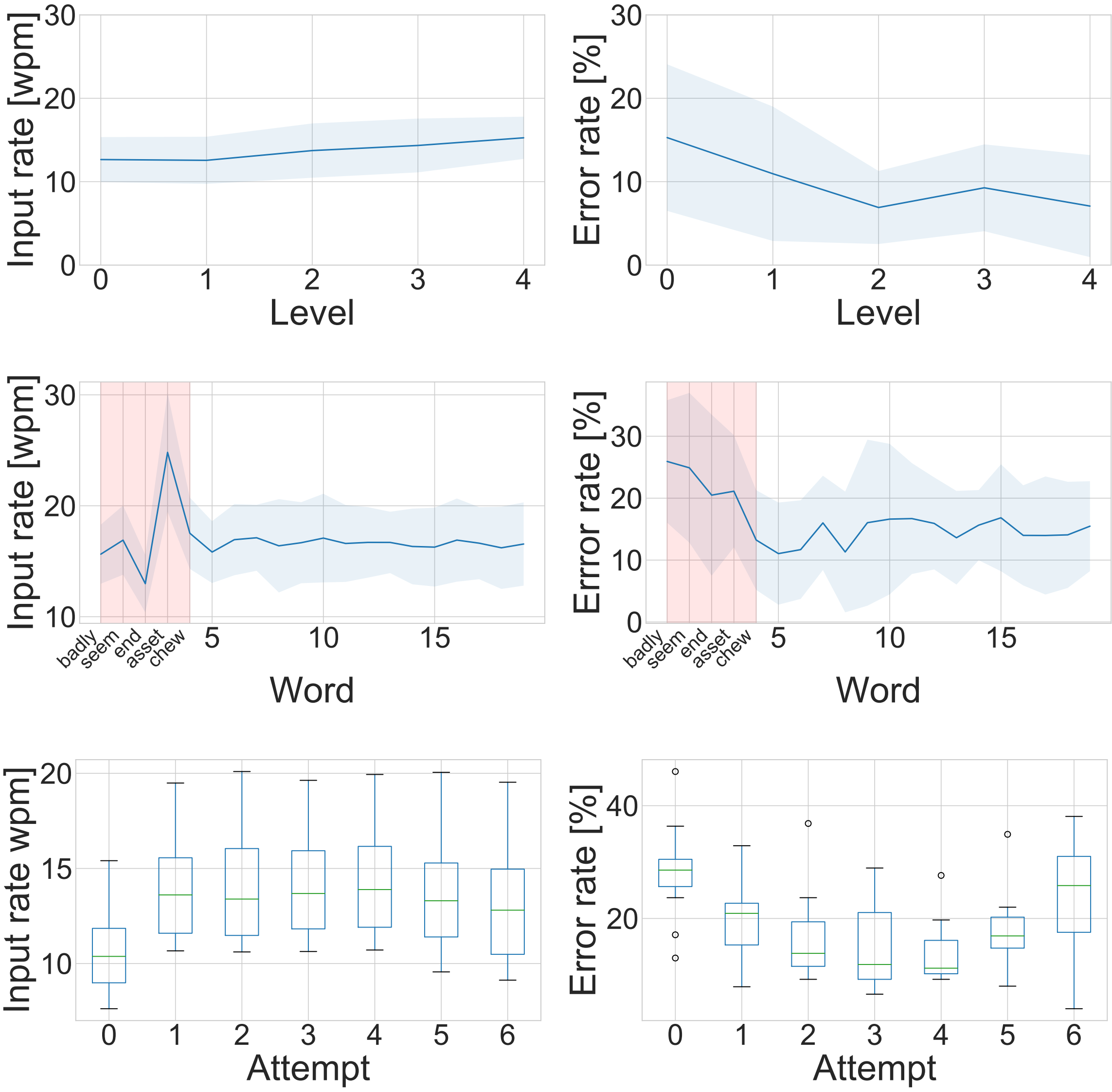

Figure 3.10: Effect of learning across LEVEL presentation, WORD presentation and ATTEMPT on the INPUT RATE and the ERROR RATE.

$14.1wpm$ for subsequent attempts. We found a significant effect of ATTEMPT on ERROR RATE ($F_{2.12,23.33} = 7.71$, $ges = 0.28$, $p < 0.01$). Pairwise comparison showed the first and last attempts did present a higher ERROR RATE than the rest of the attempts. These results revealed that participants needed to adapt to every new task. They failed on average more often on first try and were slower than the following attempts. However, it took only one try for participants to reach a stable level of performance in terms of INPUT RATE and ERROR RATE showing that the design of the block with gesture typing motions in succession did alleviate the penalty incurred by gesture recall for inexperienced users. The higher error rate for the last attempt can be explained by the fact that, if participant needed to used the seventh attempt in the block, there was a higher likelihood in proportion that it was a problematic WORD to produce. Due to the relatively small absolute difference between the first attempt and the rest, all attempts were kept for the rest of the analysis.

Outcome Effects

The influence of the testing conditions of the outcome of the task were investigated. The values of INPUT RATE and ERROR RATE as function of LEVEL are plotted in Figure 3.11 while the data containing the mean and standard deviation for those two variables is presented in Table 3.5.

Statistical analysis showed a significant main effect of LEVEL on INPUT RATE ($F_{1.53,16.80} = 108.19, ges = 0.75, p < 0.0001$). Post-hoc analysis with Bonferroni correction showed that only TP1 was significantly different than all other LEVEL. The INPUT RATE for TABLET was on average $29.4wpm$ which is in-line with what can be expected from novice users after the time of the experiments [5]. In comparison, for all OPTICAL levels, the average value of INPUT RATE was at $13.6wpm$, significantly lower. This reduction represented a drop of 54% in term of rate when participants moved from a conventional touch screen to an optically tracked surface. Our results were comparable to published data [66] where a reduction of 57% (after ten sessions) was observed between the direct and indirect input modality for their mid-air gesture technique.

Statistical analysis showed a significant main effect of LEVEL on ERROR RATE ($F_{2.85,31.34} = 13.25, ges = 0.45, p < 0.0001$). Post-hoc analysis with Bonferroni correction showed that only OP1 was significantly different than all other LEVEL. The ERROR RATE for OP1 is 26.1% on average while the ERROR RATE for the other conditions (OP2, OP4, OL2, TP1) is 9.6% on average. The lowest value for the error rate was obtained for the level TABLET with 6.2% on average. The rest of the optical condition showed a similar variance but a higher average at 11.9% for OP2, 11.9% for OP4 and 8.8% for OL2. The pairwise comparison between OP2 and OP4 with TP1, for which the differences were more pronounced than with OL2, exhibited a p-value of 0.4. These results are again comparable to published data [66] where 19.9% of the transcribed phrases did require corrections. In summary, untrained participants could produce a similar ERROR RATE while interacting with a tablet and through our system provided its interaction size was large enough.

	OP1	*OP2*	*OP4*	*OL2*	*TP1*
INPUT RATE [WPM]	13.8(3.2)	14.1(3.5)	13.2(3.0)	13.3(3.1)	**29.4(5.7)**
ERROR RATE [%]	**26.1(10.9)**	11.9(7.6)	11.7(7.8)	8.8(6.8)	6.2(6.2)

Table 3.5: Mean and standard deviation for INPUT RATE and ERROR RATE across testing conditions, with maximum values in bold print.

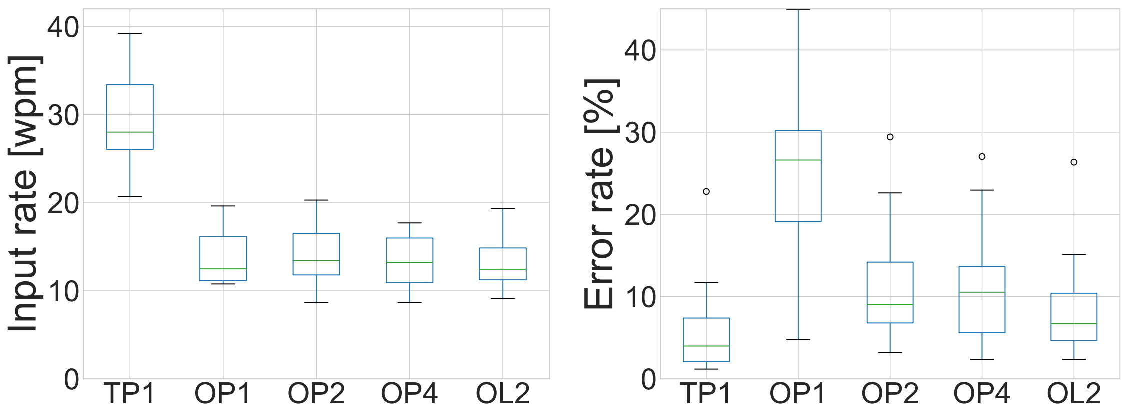

Figure 3.11: INPUT RATE and ERROR RATE across testing conditions with keyboard as reference.

3.5.7 Analysis

The result section highlighted some interesting differences in performance in terms of ERROR RATE when the participants were engaged with the smallest indirect level OP1, and in terms of INPUT RATE between TABLET and OPTICAL levels. This section will analyse further our data to understand in greater detail the reason behind these two discrepancies. In addition to the measure of results (NPUT RATE and ERROR RATE), some information can be extracted from the trace data recorded during the experiment. This data is composed of the user pointer position in the control and display space as a 2-dimensional time series. This data is more closely related to a knowledge of performance, which indicates how the participants did perform instead of what they did achieve.

Error Rate

We computed the offset at touch down as the distance between the first recorded trace data point and the centre of the key of the first letter for the target WORD. For this analysis, we selected the PORTRAIT levels only and computed the distribution of the mean value per participants (per-participant mean) as well as the overall mean value, both in control and display space measured in millimetres (mm) and pixels (px). The results are shown in Figure 3.12 with the numerical values summarised on Table 3.6. A linear regression of the display offsets with per-participants mean showed a strong linear relation in the data ($slope = 11.1, intercept = 28.0, rvalue = 0.99, pvalue = 0.03, stderr = 0.47$). Overall, the participants were not capable of maintaining the same level of accuracy in display space across the different SIZE. An average per-participant offset of $30px$ was observed on OP4, while an offset of $39px$ was observed on OP1. Due to the different CD_{gain}, the precision requirements in control space were very different. Even though participants managed to produce a $3mm$ control offset in OP1, it was not sufficient to maintain the same display performance as a $10mm$ control offset in OP4.

In an ideal scenario, participants would have been able to maintain the same level of performance across SIZE in display space. However, the linear difference in accuracy did not entirely account for the sharp increase in ERROR RATE for OP1, as compared to OP2 and OP4. By computing the offset as the distribution over all participants (without averaging), we generated an additional picture of the offset distributions as shown in Figure 3.13. The offsets for traces that generated a correct and failed decoding are plotted in blue and red, respectively. We observed a much higher likelihood for failure when the starting offset increased. The numerical values for the overall offset are presented in Table 3.6 in the last column. For the PORTRAIT orientation, the keyboard has a demi-key width and height of $50px$ and $75px$, respectively. As a result, for level OP1, the starting offset was, on average, $3px$ greater than the width of the intended key. The much higher level of failure for OP1 should be, in part, attributed to the decoding algorithm which seemed to rely heavily on the correctness of the first selected letter to produce a correct decoding.

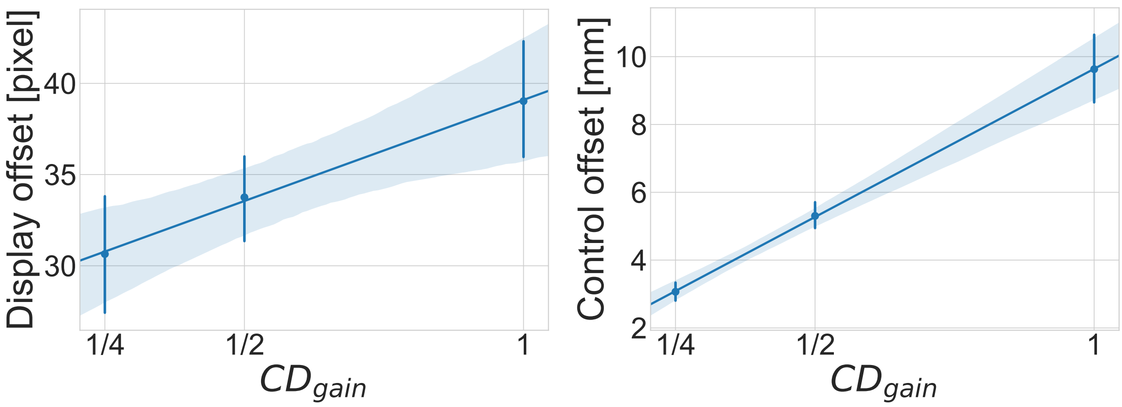

Figure 3.12: Offset on the target acquisition of the first letter of a target word in display and control space on the left and right, respectively.

		Per-participant mean		Overall
		control touch	*display touch*	*display touch*
LEVEL	CD_{gain}	*offset [mm]*	*offset[px]*	*offset[px]*
OP1	1	3.07 (0.49)	**39.03 (6.18)**	**53.2 (63.2)**
OP2	1/2	5.30 (0.70)	33.75 (4.43)	44.8 (59.0)
OP4	1/4	**9.63 (1.87)**	30.64 (5.95)	41.0 (56.6)
OL2	1/1.7	5.52 (0.83)	41.41 (6.25)	57.1 (87.9)

Table 3.6: Touch down offset for OPTICAL computed in display and control space, averaged per-participants on the left, with maximum values in bold print.

In addition, the starting touch offset might not be the only factor explaining failure or success

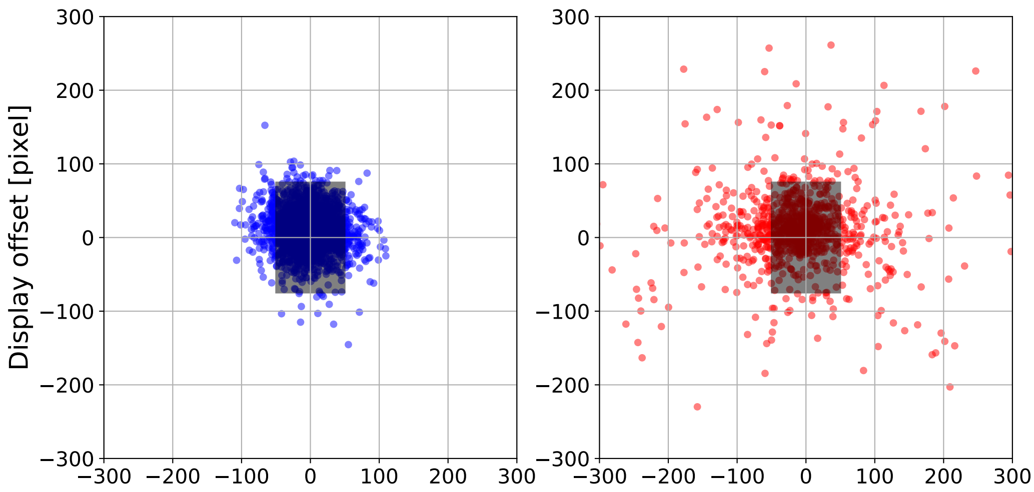

Figure 3.13: Display offset across PORTRAIT levels for successful and failed attempts in blue and red, respectively. The shaded area represents the key boundaries with dimension 100 by 150 pixels.

of the task; some attempts that did not present a great offset still induced a task failure, as pictured by the red points within the key boundary in Figure 3.13. Therefore, it is likely that the poorer accuracy exhibited for smaller SIZE, in terms of starting offset, carried through the rest of the gesture with more subsequent inaccuracies, and as a result induced more decoding errors than for OP2 and OP4.

Input Rate

The results for the INPUT RATE did show a statistically significant difference between TABLET and OPTICAL, but not between OPTICAL levels. Since we did not collect the trace data for the condition on the tablet, the analysis was focused on OPTICAL levels. For the trace data, we computed the first, second and third derivative of motion representing velocity, acceleration and jerk, respectively. We used a Savitzy-Golay filter, see [77] for reference, with 15 sample window length and a $3rd$ order polynomial. The reported results for speed are computed as the norm of velocity. For the subsequent analysis, only PORTRAIT levels are considered to allow for a fair comparison, but results for LANDSCAPE are included for reference.

The results for the speed in control and display space are presented in Table 3.7. A statistical analysis showed an effect of SIZE on display speed albeit with a small effect and explained variance ($F_{1.78,19.53} = 4.61$, $ges = 0.08$, $p = 0.03$). OP1 was the condition with the greatest average display speed with a mean at $457.6px/s$, similar to OP2 with a mean speed of

$450.4px/s$, while OP4 presented the lowest speed at $388.5px/s$ equivalent to a reduction of 15%. As suggested by the equal INPUT RATE, the control speed showed a great range of values with an average value of $3.6cm/s$ for OP1, to be compared to $12.2cm/s$ on average for OP4. In other words, participants were increasing their average hand speed by a factor 3.4 between the conditions OP1 and OP4. Also, the pointer display in OL2 was on average comparatively higher than all other levels, with a mean value of $599.8px/s$. This result will be focused on in the discussion.

LEVEL	CD_{gain}	*pointer control speed[cm/s]*	*pointer display speed[px/s]*
OP1	1	3.6 (0.9)	**457.6 (114.9)**
OP2	1/2	7.1 (1.7)	450.4 (110.7)
OP4	1/4	**12.2 (2.9)**	388.5 (93.4)*
OL2	1/1.7	8.0 (1.9)	599.8 (145.0)

Table 3.7: Pointer speed for OPTICAL computed in display and control space, with maximum values in bold print and statistical significance marked by asterisk.

Next, the derivative profile of the traces produced by the participants across the different levels of PORTRAIT were investigated. The temporal evolution of the user pointer has been previously modelled, especially in pointing task. Recently, Muller et al. [78] have investigated different models for such task, revealing that differences in user behaviour dependent on the index of difficulty.

To gain an insight into these temporal behaviours, the number of zero-crossings of the velocity, acceleration and jerk were computed, refer to Table 3.8 for numerical values. These measures can be indicative of different steering behaviours. For instance, a zero-crossing in the speed profile indicates a change of direction which could reflect an overshooting in a target acquisition task. A statistical analysis showed a significant effect of SIZE of number of zero-crossings of velocity ($F_{1.38,15.23} = 8.53$, $ges = 0.18$, $p = 0.006$). Post-hoc analysis showed that OP1 presented on average more zero-crossings than OP2 and OP4. No statistical difference was found for the zero-crossings in acceleration and jerk, albeit OP1 did exhibit a greater value for both measures.

# of zero-crossings	*OP1*	*OP2*	*OP4*	*OL2*
velocity	**8.7 (1.7)****	7.3 (1.6)	6.9 (2.0)	7.7 (1.6)
acceleration	**11.9 (2.2)**	10.6 (2.2)	10.5 (2.7)	11.2 (2.3)
jerk	**21.8 (5.8)**	19.1 (5.6)	19.4 (6.9)	20.8 (6.1)

Table 3.8: Mean and standard deviation for the number of zero-crossings of speed, acceleration and jerk across testing conditions, maximum value in bold print and statistical significance marked by asterisk.

The signification of an increase in zero-crossing of the velocity is illustrated on Figure 3.14 where the speed profile is plotted for the word BADLY in level OP1 and OP4, represented in blue and orange respectively. We can observe a much lower minimum velocity for OP1, while in OP4 the speed is maintained to a higher level throughout.

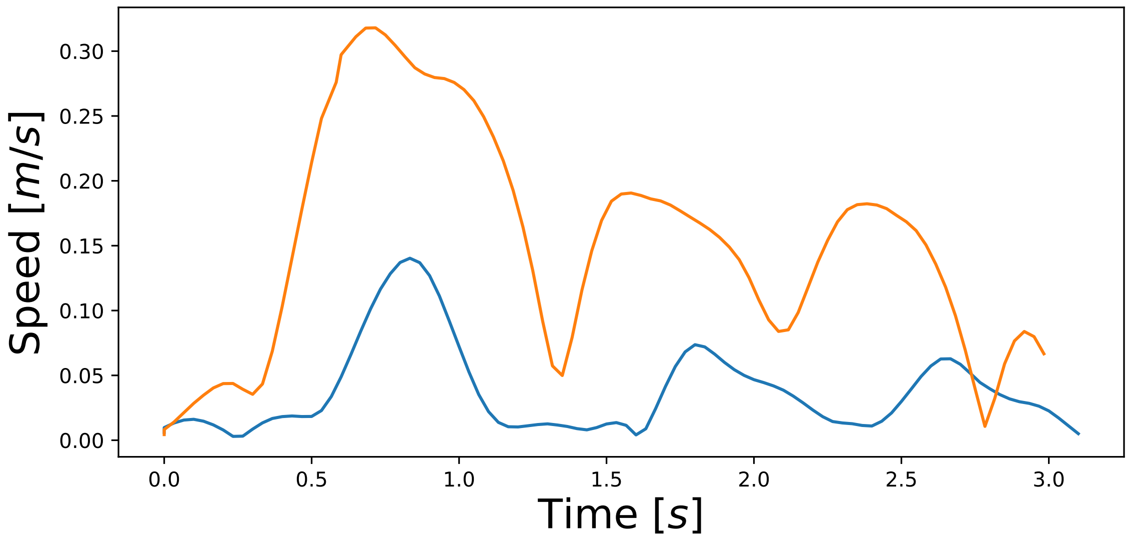

Figure 3.14: Speed profile for WORD BADLY in level OP1 and level OP4 in blue and orange, respectively. The differences between OP1 and OP4 appear in the dynamics of the motions, where fewer zero-crossings in velocity are observed for OP4.

3.5.8 Qualitative Data

The data from the NASA-TLX showed that for all polling categories, TABLET was the least demanding condition, followed by OP2 in OPTICAL. OP1 was associated with the highest *mental demand* and *temporal demand*, while OP4 scored highest in *physical demand* and *effort*. A statistical analysis showed significance for *physical demand*, *effort* and *frustration* but no effect on *mental demand*, *temporal demand* or *performance*. The big picture that emerged from this data was that OP1 and OP4 were requiring more efforts from the participants as well as induced more frustration, and that OP4 in particular was particularly demanding in terms of physical requirements. The data collected through the NASA-TLX is summarised in Table 3.9.

In addition to the NASA TLX data, participants were also asked to rank the different levels according to their preference with #1 as the least preferred and #5 the most preferred level. The level on the device TABLET was the most preferred for all but one participant. The second preferred condition were the levels with SIZE 2 with OP2 and OL2 ex aequo, followed by OP1 and finally OP4. OP4 was designated as the least favourite for all the participants. The data is presented in the Table 3.10.

LEVEL	*mental demand*	*physical demand*	*temporal demand*	*performance*	*effort*	*frustration*
OP1	10.3 (5.0)	9.4(4.6)	9.1(4.6)	7.3(3.6)	11.7(4.4)*	10.3(3.9)*
OP2	**7.7(3.9)**	**9.3(4.9)**	**6.9(4.0)**	**5.4(2.9)**	**8.7(4.6)**	**6.4(4.0)**
OP4	9.3(3.0)	13.3(4.1)*	8.9(3.5)	7.0(3.5)	12.8(3.2)*	10.1(3.9)*
OL2	8.8(3.5)	9.7(3.6)	8.3(3.2)	6.8(4.1)	9.8(4.1)	8.0(3.8)
T	*4.6(2.6)*	*3.8(2.7)*	*5.8(3.9)*	*3.2(1.8)*	*4.2(3.0)*	*3.7(3.1)*

Table 3.9: NASA TLX data with the minimum value in bold print and statistical significance marked by asterisk.

LEVEL	*preference ranking*
OP1	2.3(1.5)
OP2	**2.8(1.2)**
OP4	1.0(0.7)
OL2	**2.8(1.5)**
T	*4.9(0.3)*

Table 3.10: Data from the preference ranking, with the minimum value in bold print and statistical significance marked by asterisk.

3.5.9 Discussion

The data presented in this chapter strongly suggests that virtual surfaces created through optical tracking are suitable for ecologically valid tasks such as gesture typing. The results obtained are, indeed, interesting and in comparison, with a conventional interaction, they showed that the ERROR RATE can be similar even though the INPUT RATE was reduced by half. The results from other related research studies [52, 66, 67] can be used to put our findings in perspective. These studies have similar findings even though they investigated different tasks or used a different sensing technology.

Concerning the ERROR RATE, we have looked into explanations for the much higher values for OP1. It is interesting to point out that Vertanen et al. [67] did find a similar effect in their investigation. Beyond the apparent limitation in the user motor control, the proportion played by the decoding algorithm was however untractable at this point.

With regards to the difference in INPUT RATE between OPTICAL and TABLET, in [66], Markussen et al. have a discussion about the apparent slowness of their similarly indirect input technique. They mention that the decoupling between control and display space is mentally taxing for the users, according to the principle of stimulus-response compatibility [46]. They also point out that the reliance on the visual feedback induces a penalty which translate into a slower movement speed as compared to a direct interaction. Comparing the INPUT RATE in OP1 and TP1 shows that participants were indeed moving on average slower in the indirect conditions. We have also wondered why the participants were slower and after investigation it appeared that the display presented a non-negligible delay in the order of $100ms$. Some recent work [79] has highlighted this issue (which can potentially be overcome through software methods [80]). However, we have also observed different pointer display speeds, most notably in LANDSCAPE mode where the usable display area was bigger. This strongly indicate that the main effect bounding the user performance is not only a delay in the feedback.

The lack of effect on INPUT RATE in OPTICAL was not expected, as scale should be a basic component of performance [53]. Based on the effect shown in [52, 49], we did expect to measure an effect of scale. We know that for hard tasks, the performance should present a U-curve and given the lower INPUT RATE in OPTICAL, it appears that the task can be deemed hard enough for the participants. One potential explanation lay in the fact that the task we employed was more complex than pointing or circular steering and the measurements we made were in return noisier. The scale that were chosen in these two studies are also different. Accot reports ranges from 1 to 16, while we used 1 to 4. The range chosen was equivalent to a paper size of A5 to A3 which we deemed sufficient for a plausible interaction scenario. Indeed, users did indicate that OP4 was bigger than wished through their preference rating and qualitative feedback.

One significant effect of scale we measured was the induced differences in more intrinsic properties of the data, namely in the distribution of derivatives of the participants motions. Some work has been conducted on the dynamics of pointing task by Muller et al. [78], where modelling on the time series of a 1-dimensional pointing task were carried out. Adapting such models to gesture typing would be needed to provide a further insight into our data.

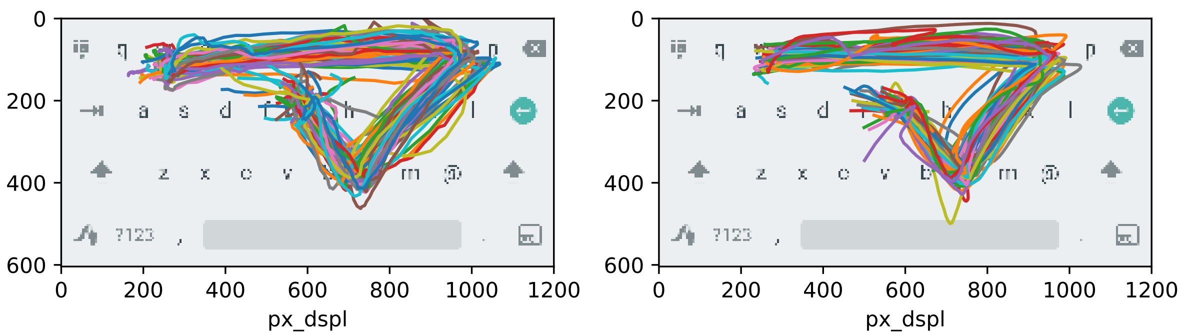

Figure 3.15: Plot of traces in display space for successful attempts of word “wrong” for condition OP1 and OP4 on left and right, respectively.

Finally, we were also interested in the stability of the traces, potentially indicating differences in term of user behaviour and performance. Figure 3.15 shows the position data for OP1 and OP4 when the target word is “wrong” and the task successful. We can see the bigger distribution on touch down around the letter W for OP1, and also distinguish some differences in term of noise in the traces, with OP1 presenting less smooth trajectories and potentially covering a bigger region of the keyboard. This idea will be investigated in the last chapter.

3.6 Conclusion

In this chapter, we have proposed a novel upper limb interaction which affords gesture typing through virtual surfaces created by means of marker-less optical tracking.

The research questions we set out to investigate were:

- **RQ0.1**: How computational models can be used to transform image data into a material enabling touch interactions?
- **RQ0.2**: How does this new interaction compare to the control condition of a tablet with capacitive touchscreen?
- **RQ0.3**: What is the influence of size on the text-input performance and on the interaction perceived quality?

RQ0.1 was addressed by the description of our tracking pipeline with a 3-steps process which consisted in the signal processing of depth images to segmented pointclouds, followed by a regression task for estimating the pointer position and a classification task for estimating its touch property. We have shown that the regression can be improved with the addition of a Kalman filtering stage, which reduced variance in positional errors in a steering task. We have also shown that a model based on neural networks produces competitive results for the classification tasks with an AUC of 96%.

A user study was designed to answer **RQ0.2** and **RQ0.3**. We were interested in comparing our novel interaction with the control condition of a tablet interaction. We have measured the input rate and the error rate on the control condition and on our tracker for different levels of input sizes. We have shown that the input rate is halved when the interaction happens on the tracker, as compared to the reference input rate measured on the tablet. A result similar to what has been obtained in a study on mid-air gesture typing. We have also shown that the error rate achieved with the optical tracker is not different to the one obtained on the tablet, provided the input size is big enough.

We have also recorded the effect of scale on the interaction. Investigating the reasons for a higher error rate on the smallest level, we found out that accuracy levels were not maintained throughout the change in input size. To be precise, the offset on the first touch point was on average higher than the half-key width indicating a recurrent mis-acquisition of the first letter of the word. We have also shown that the user dynamic behaviour was affected by the change in input size. The number of zero-crossings were significantly higher for the smallest size. This result shows that some of the effects of scale could only be measured on intrinsic parameters motions, and not measures of outcome. The interaction quality has also been measured and participants voiced a strong preference for the control condition, followed by the intermediate input size which produced limited errors and did not require extensive arm motions. A strong dislike of the biggest input size was expressed by the participants, it was ranked last.

Chapter 4

Rehabilitation Through Common Gameplay

Summary. This chapter proposes a novel interaction for the rehabilitation of reaching capabilities of the upper limb for users with spinal cord injuries, stemming from the interaction of gesture typing through virtual surfaces, which presents similarities in terms of user motions. A series of design workshops is used to understand the needs of occupational therapists for gamified interactions and leads to the creation of an input control modality that interfaces with off-the-shelves video games. A user study is designed to understand how parameters of the user interaction loop impact the overall user performance. It shows that it is possible to maintain the user performance by altering the game framerate and that rehabilitation goals can be met through an optimisation of the game controller. We then argue that a model of user behaviour is key to afford an enjoyable user experience. A computational approach is used to build a probabilistic model of user behaviour from reference gameplay sessions. The probabilistic model provides in this context a low-latency measure of performance that is essential to inform the optimisation process.

4.1 Introduction

A collaboration with a team of occupational therapists from the Queen Elizabeth University (QEU) Hospital of Glasgow started during a workshop meeting organised within the frame of the European project Moregrasp, see section 2.1. A demonstration of the system, developed in Chapter 3, was made to the participants and prompted questions about possible alternative use-cases for users with spinal cord injuries. Occupational therapists identified a similarity between gestures needed for text-input and motions needed for the functional rehabilitation of the upper arm, adding that common exercises performed in that context were repetitive by nature [81] and sometimes unnecessarily "dull, tedious and boring". The discussion which followed focused on potential ways to remedy their patients' motivational issues during upper arm rehabilitation.

A common strategy in the field of HCI for fostering engagement is to leverage patients' natural propensity for play, and the process of *Gamification* has in that regards been the most widely employed approach. *Gamification*, as defined by Deterding et al. [82], is the process of incorporating "video game elements in non-gaming systems to improve user experience and user engagement", which in practise has often relied on adding badges, points and rewards systems analogous to what is present in common games, see [83, 84] for a review. The field around *Gamification* has grown increasingly complex and includes different flavours and approaches; sometimes intertwined but tackling the same problem. The creation of games, by definition in opposition to the concept of *Gamification*, has been proposed with *Serious Gaming* and *Games with a Purpose*. These two approaches rely on the design of activities similar to games with the goal of enacting a serious activity, see [85] for a critical review. In particular, *Exergaming* is a subcategory of *Gamification* whereby physical activity and/or exertion is elicited. The definitions of and relations between these four different categories is subject to fluctuation and debate, however a potential superset of these categories comes with *Playification*, defined by Nicholson et al. [86], which focuses on the broader element of play in the activity.

Recognising the value of video games in driving user motivation and engagement, we propose that the activity that serves as a motivator should be as close as possible to original digital games and that, if possible, already made games should be used instead of purposefully created inspired reproductions. There are some obvious advantages to this approach: engaging digital games are valuable artefacts requiring careful design and access to resources few can afford; successful digital games sometimes transcend demographics and their familiarity to users would make rehabilitation instructions and goals easy to understand. However, users ought to be induced into their rehabilitation via the game's own gameplay. The interaction with the game should require users to perform the motions predefined by the occupational therapists. These are however unlikely to match those needed by default from the original unmodified game. Alterations to the player's interaction loop will be required, so that physical exercises become a by-product of play.

In the field of HCI, some recent research has followed the same principles. Walther-Franks et al. [87] have proposed to use off-the-shelf games for generating motivation in exercising. They derived a 4 step process through which such games can be adapted. It comprises the choice of the game, the creation of a control overlay, the design of a feedback overlay and the adaptation of workouts. In particular, they mention the challenge of "finding mappings from control input to game-action" and take a design approach to solve this problem. Using a character-based action game, they find 1-to-1 mapping between the avatar and the user body.

Ketcheson et al. [88] pushed the idea further and adapted two well-known games (Half-Life and Skyrim) for exercising. They identified three type of approaches for converting off-the-shelves games: black box conversion, source code modification and modding. Each relating to the degree to which the original game can be modified. In their work, they used games where the player's avatar is powered by pedalling a stationary bicycle with the target heart rate as their exercising goal measured through Borg rate [89]. In addition, they used modding to change the game dynamics and provide power-ups to users in order to balance the game difficulty. Experiences of their users were collected through the IMI scale [90] which provides an access the participants self-report measure of intrinsic motivation. What can be taken away from these two attempts, is that there is an important place accorded to design, especially the design of a new control modality, even if the goal is to use already existing games. Also, user experiences were measured through questionnaires which, despite being detailed and informative, present the very salient drawback of presenting results after the experiments with a considerable time delay.

Rehabilitation research related to the task of upper limb rehabilitation for stroke patients, who present similar symptoms to patients with spinal cord injuries, has also used off-the-shelves games. Already in 1993, Sietsema et al. [91] investigated the effect of using a Simon game for reach rehabilitation and showed an improvement in movement amplitude when using the game as compared to traditional exercises. More recently, the game Fruit Ninja and the sensor Kinect have been used in conjunction [92, 93], producing as well positive results in clinical tests after prolonged use. Game difficulty, via modding, was adapted in [93] to the new interaction paradigm and target audience. These findings are similar to what can be learned from a more medically oriented field. Barret et al. [94] have reviewed the topic of upper limb stroke rehabilitation and, even though they support the *Gamification* approach, through the creation of new games, their recommendations can be understood in a broader context and used with off-the-shelf games. They point out that goals and rewards are important tools for motivation, just as well as challenge and difficulty, and that meaningful play and feedback need to be provided. The six main features of importance, selected after the frequency of occurrence in their review of published survey papers, taxonomies and frameworks, were:

- Socialisation
- Motivational feedback
- Simple interface
- Challenge
- Appropriate cognitive challenge

- Adaptability to motor skill level

The first five features are not exclusive properties of serious games or gamified activities and can be easily found in off-the-shelves games through a conscious selection. However, *Adaptability to motor skill level*, and as we will see *Challenge*, points toward the main problem with unmodified games: adaptability.

In this chapter, we are setting out to propose a novel upper arm interaction for the task of rehabilitation of patients with SCI, which also implies the design of a new control modality. The context and activities that were carried out with OTs to inform the design of such interaction are first described, before a computational approach is formalised. The result from a user study aiming at measuring the effect of the new control modality are presented before moving into a modelling step targeting the optimisation of our design.

4.2 Context

Following the initial workshop meeting with the team of occupational therapists, various design activities were carried out to understand the need of medical practitioners and patients. We engaged in workshops at the University of Glasgow and at the QEU hospital to learn about the activity of game design for rehabilitation and understand in more depth the daily work of occupational therapists, respectively.

4.2.1 VR Design Workshop

A 2-day design workshop organised within the University of Glasgow focused on the use of Virtual Reality (VR) for neuromotor rehabilitation. This workshop gathered of a wide range of profiles among the attendees including design and game practitioners, PhD students from various schools and stroke survivors. The workshop was led by an alumni from the University of Glasgow school of medicine who practised as an NHS emergency medicine consultant. She was assisted by a well-known international emerging medical technology adoption and regulatory expert. The attendees were divided in small homogeneous groups, ensuring that at least one stroke survivor was present in each groups. The goal of the workshop was to bring together a diverse group of participants with common interests to have a discussion about the use of VR technology in neuromotor rehabilitation and engage in a design activity.

The first session of the workshop was designed to give attendees some background of game design. A simple motivational model describing four types of players was outlined, with the *killer* motivated by competition, the *master* driven by skills learning and acquisition, the *social* whose primary interest is interaction with others virtual or non-virtual agents and the *explorer* whose curiosity motivates play. We also shared experiences to other members of the group: stroke survivors talked about their experiences of living through a daily rehabilitation routine. They mentioned the need for feedback on the progress as an intrinsic motor for motivation and pointed out that individual rehabilitation routine are different from one person to another and need to be tunable in terms of difficulty depending on timely needs. They also stressed that motivation came often from the group itself: as different patients struggle through rehabilitation, the support provided by others was a key factor to continue engaging in tedious recovery. On the other hand, VR specialists detailed the possibilities offered by the current state of the technology. We also talked about gaming habits and a group discussion about what games participants knew, played and enjoyed singled out Arcade games as a recurring category.

The second session was aiming at refining the ideas from the first session and designing an envision scenario. We created a paper prototype representing a VR rehabilitation session that was latter presented to the workshop attendees. The prototype told the following story: a user starts a rehabilitation session by wearing the VR helmet and is welcomed to an interface that let her chose a game, an exercise or access her profile. The game menu grants access to a list of Arcade games (Ice Hockey, Grand National, Simon or Drum Hero) with details about the body region each game would stimulate. Similarly, the profile menu let the user chose a body region to stimulate and access the list of compatible games, allowing the opposite operation to the one afforded by the game menu. The user menu could also present the information stored on purpose by her therapist and let the user access recommended exercises. For each workout, a summary of the performance is presented in terms of score, but also includes more detailed analysis, such as left or right arm involvement, accuracy or the evolution of the user performance over time.

The lessons learned for this workshop were in agreement with the summary by Barret et al. [94] reported earlier: the need for socialisation as motivation force, the importance of feedback about the progress made or the reliance on fair challenge when playing a game were all mentioned. It also made apparent that satisfying all needs with one system is a challenge. For example, socialisation implies group play which could be addressed with competitive play or collaborative play. However, competitive play poses the problem of handicap balance between players with different limitations and collaborative play makes the choice of games greatly reduced. One idea that I decided to keep forward was the use of Arcade as a seed to design. Arcade games appeared familiar to the participant who had suffered a stroke in our group, and since Arcade games usually include a succession of different short stages they proved easy to repurpose as individual exercising sessions in the paper prototype.

4.2.2 Workshop with Occupational Therapists

Following the VR workshop, a collaborative activity was organised in the QEU hospital.

Rehabilitation Exercises for Arm Reach

The first goal of the workshop was to understand the routines used for the rehabilitation of the reach of the upper limb of patients with spinal cord injuries. The target group for the OTs were patients with high level of injury, typically at levels C5 and C6, for which the arm functions are severely compromised, see section 2.1. The OTs used a functional approach to the rehabilitation aiming at restoring motions that could be used for everyday tasks such as brushing teeth, taking care of one's hygiene or eating; activities mostly focused on the upper arm. They described a range of games they used to elicit motions needed for this tasks. Some games necessitated fine motor control of the hand, such as the "labyrinth marble" game [95] which requires players to control two small knobs in rotation which in turn tilt a surface supporting a labyrinth that a metallic marble is supposed to navigate. The motion targeted were small rotations of the forearms. The "wire loop" game [96] was also employed. It requires players to hold a rim and to move it along a metallic wire without ever getting in contact with wire. Depending on the shape of the wire, which was malleable enough to be deformed on purpose, different motions recruiting the elbow and the shoulder were elicited.

For earlier stage of the rehabilitation however, when fine motor control of the finger is absent from the patients, two artefacts were primarily used. The first one was a skateboard, which sketch is reproduced on Figure 4.1, onto which the arm of the patients is strapped and whose purpose is support to the user's arm while a back and forth motion is produced, helped by the reduced friction produced by the skateboard's wheels. The instructions from the OTs were for the patients to "motion towards the window" or to "reach forward" with continuous feedback such as "keep breathing", "relax" or "stretch". The issues encountered were that the wheels under the skateboard were likely to fall from the edges of table and that the lack of intrinsic purpose was quite apparent to their patients. The other artefact used was a plastic roll that needed to be moved along one direction. A Velcro tape was attached to the roll and to the table in order to increase their friction. As a result, more force was needed to be applied by the patients to move the roll along. Here again, the simplicity/artificiality of the task was impacting its overall purpose. These exercises did describe the motions that occupational therapists were expecting from their patients: motions of the upper arm on a planar surface with different directions and ranges with a friction potentially altered.

The occupational therapists did have access to video games systems (Wii and Playstation) but declared having a seldom use of them since, even for simple games, some the motions required for interaction could not be produced by their patients. No other digital artefacts

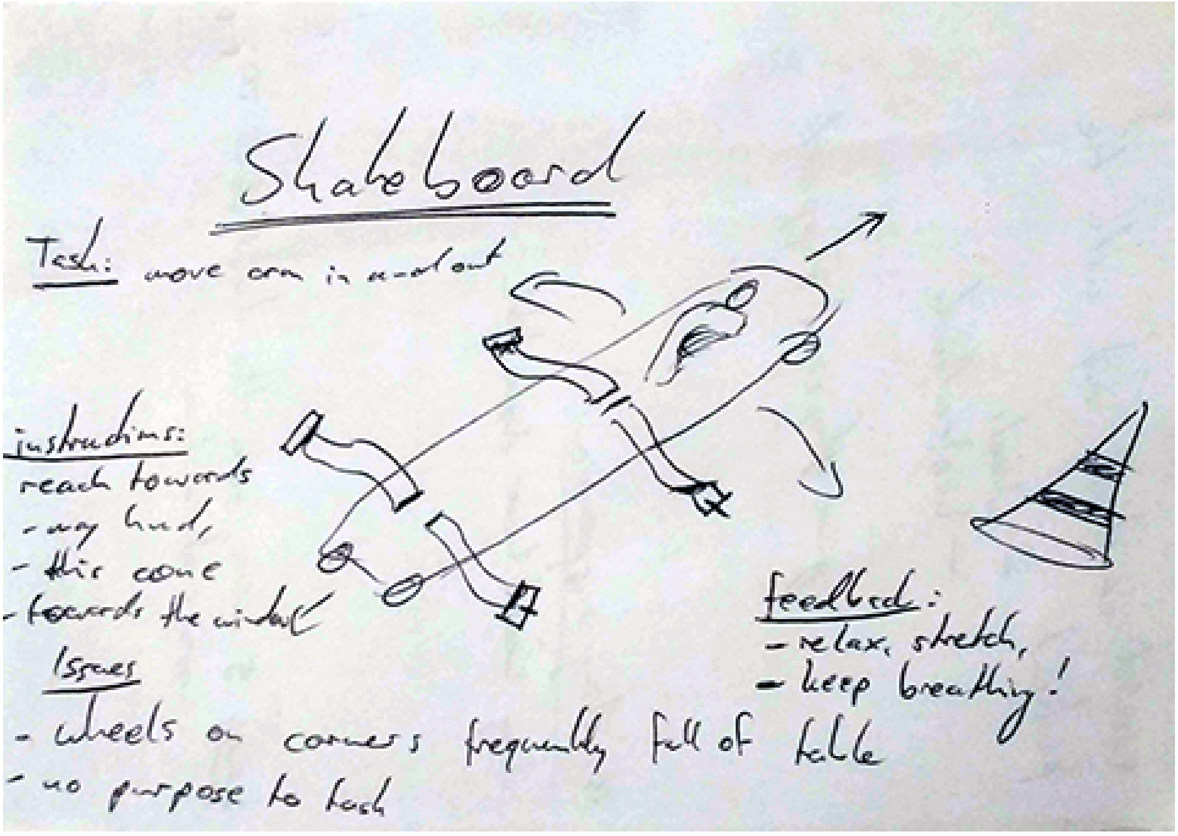

Figure 4.1: Sketch on the "skateboard" drawn during the workshop. A rectangular surface was supported by four rotating wheels which allowed movements in all directions. It was also equipped with a pair of straps to ensure that the user's arm was kept in place. Additionally, a handle was positioned at one end of the rectangular plate allowing the user to grab onto the device and help for its control.

were employed in the rehabilitation process.

Co-design

The output of the VR workshop was presented to the OTs in a final workshop. The goal was to collaboratively define an artefact and an interaction for the realisation of an engaging upper limb rehabilitation through reach exercises of users with high SCI. In particular, the question of matching motions that occupational therapists needed for rehabilitation with those elicited from a range of potential Arcade games was central. It addressed the challenge of "finding mappings from control input to game-action" pointed out by Walther-Franks [87].

We first engaged in a brainstorming session were a list of candidates for the games and a list of rehabilitation motions were proposed. Games such as Whack-a-Mole, Bubble Witch, Pong or Frogger [97] were mentioned from the instruction of focusing on Arcade games. The controls needed to implement their gameplay was then discussed. While some required four ways directional actions, for aiming at a target for example, others only required button presses for triggering specific actions. A game such as Frogger would only require four separate actions which could be executed single-handedly. It was noted that there exists a breadth of control schemes and that to simplify the workshop, games with simple controls should be focused on: they would be easier to match with rehabilitation exercises due to their comparatively reduced complexity. It also became apparent that to interface with several different games, the design of a simple control modality could be sufficient provided their control scheme was compatible. The OTs did not have a strong opinion on the choice of specific games, but emphasised the possibility of choosing different games depending on the occasion.

The second discussion was targeting the different exercises the OTs would like to see implemented by the games gameplay. The motions that patients were producing when interacting with the skateboard, mentioned earlier, were used as a reference. These were long reaching longitudinal movements which were compared to target acquisitions, similar to Fitts' experiments. This led into the matching exercise, where the motions described for game control were pictured on the tabletop where rehabilitation usually took place. Since the instructions from the OTs were to perform long reaching motions with the arm in contact with a planar surface, and that they were encouraging their patients to reach always further, the idea of placing actionable controls on different location on the surface was proposed.

In summary, the OTs expressed some specific requirements. The interaction should happen on a tabletop and present the same properties in terms of motion as the original exercises proposed by the occupational therapists. The difficulty of the task should be adjustable based on the current patients capacities. It should be possible to record and analyse the performance of a given patient after a session.

From an HCI perspective, we also expressed one requirement. The game used for motivation purposes should be unmodified, or as close to its original form as possible, to harness its natural capacity at engaging players. One way to achieve this while being able to control the game difficulty is to play on the flow of time in the game space. This has the advantage of preserving the game visuals and the game mechanics. We argue that this does not alter the game too much, especially for arcade games, and is indeed a popular game design. Think about Tetris, where the flow of time is constantly increasing as the levels go; or Bullet time in current popular video games such as Max Payne[1] or Red Dead Redemption[2], which slows down the flow of time to enable easier target acquisitions of enemies. Expressed from an information theoretic viewpoint, changing the flow of time is equivalent to reducing the information throughput required by the game for interaction. The information throughput, in bit/s, is correlated with the number of actions per unit of time. A slow down in the game

[1] http://www.rockstargames.com/maxpayne/
[2] http://www.rockstargames.com/reddeadredemption

time flow will reduce the number of actions needed from the user, whose flow of time is unaltered, for maintaining the same information throughput.

This last workshop enabled us to have enough information to propose an interaction.

4.3 Proposed Interaction

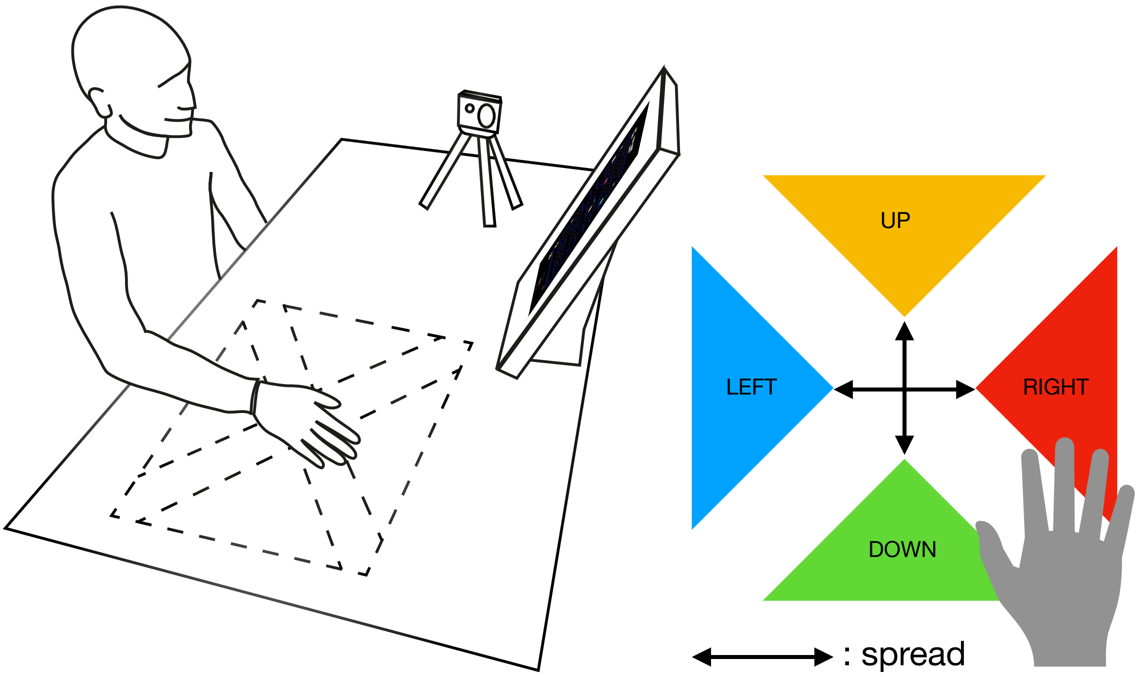

Figure 4.2: Sketch of the gamepad interaction. A user controls a digital game through 4 actionable areas placed on the tabletop. Optical tracking creates the 4 virtual controls and follows the user's hand (left). The virtual gamepad, solely defined by the parameter *spread*, maps hand positions to the four commands UP, RIGHT, DOWN and LEFT (right).

The design that was agreed upon for the reach rehabilitation of the upper limb of users with a spinal cord injury is sketched on Figure 4.2. A user is seated by a tabletop on top of which reaching motions are performed. An optical device serves the dual purpose of tracking the user's hand and creating a virtual surface which affords 4 control buttons arranged in a cross shaped fashion on the planar surface. Each of the buttons is activated when the hand of the user is detected within their control boundaries. Buttons are mapped to directional commands (UP, RIGHT, DOWN, LEFT) equivalent to the arrow keys of a keyboard. When buttons are triggered, actions are sent to a computer that runs a digital game. The computer produces an audio feedback that indicates when controls are registered, on top of the common game audiovisual feedback that is shown on the screen facing the user and conveyed through headphones. In order to partake in the gameplay, users issue commands in succession by motioning their hand to the specific locations of the virtual controls. As for the game, *Pac-Man* [98], one of the most iconic arcade games was chosen. The controls are thus connected to the four action that the avatar in *Pac-Man* can perform: turning up, left, right and down.

4.3.1 Reformulation of the problem

Given the proposed interaction, a computational view of this interaction can be formalised.

Mapping Motions & Adapting Difficulty

The motions used for game control and the one targeted for physical exercising are unlikely to be identical. When both motion sets are aligned, exercises become a by-product of play. In most cases, a mapping between motion sets needs to be created, usually through interaction design with the *making of a new input control modality*. Traditional hand-held controllers have been replaced or augmented with motions capture systems [99], training apparatuses [88]. These sense additional parts of the user's body the physical rehabilitation seeks to solicit, that can also be used for game interaction. Yet, such alterations to the player's interaction loop have potential negative consequences on performance: different input limbs afford different information throughput [51]. Using arms or legs instead of fingers to interact with a game will invariably wield inferior control. This has been repeatedly reported as a source of frustration when the newly afforded levels of control are not sufficient for enacting the originally intended gameplay [100].

Providing a reasonable challenge to the user is key to an engaging experience. As phrased by Przybylski et al. [101], the "mastery of controls plays an important role in game motivation, largely as a necessary, but not sufficient, condition for achieving psychologically need-satisfying play." For off-the-shelf games, adapting difficulty, when implemented, has been achieved through modifications to games: additional power-ups have been used [88] or game content was removed [93]. Tuning these modifications is most often done empirically, showing that setting their balance on gameplay remains a challenge. Additionally, physical exercises require some degree of parametrisation, as voiced in the literature related to rehabilitation [94]. Intensity or amplitude should be variable and specified by training goals, configurable by therapists. Exercises should also be adaptable to patients based on their individual capabilities, different users groups (age, injury) having different needs.

Designing an interaction that balances these objectives - a new control that elicit desired motions while providing enough control to enact gameplay and sufficient flexibility to modulate exercise intensity - is a complex challenge.

Computational Design

Given the complex nature of the design task, we resolve to explore computational methods. Computational design is defined by considering design activities as an optimisation problem, where design constraints can be expressed as objective functions in the space of possible designs [62]. Under this description, the controller and the game designs are the variables which can be altered so that rehabilitation and engagement objectives are fulfilled. Formulated as such, a design variable $\mathbf{x}$ collects what can be changed in the interaction loop:

$$\mathbf{x} = [\theta_{control}, \theta_{game}] \in \chi \tag{4.1}$$

where χ represents a multi-dimensional design space which describes all the possible designs for the game controller, and all the possible alterations for the games. The goal is to find values for the parameters that minimise a cost function g:

$$\mathbf{x}^* = \arg\min_{x \in \chi} g(\mathbf{x}) \tag{4.2}$$

where the value for the cost function is minimal as rehabilitation goals are met and user engagement is maintained. Note that an alternative but equivalent view could use an objective function h which value increases as goals are met. The goal for the optimisation becomes then:

$$\mathbf{x}^* = \arg\max_{x \in \chi} h(\mathbf{x}) \tag{4.3}$$

In this work, and in contrast with the literature, the objective function is expressed as a distance to an unmodified gameplay. In other words, given unmodified gameplay provides engagement, we are interested in finding solutions where the difference between what is observed and what was produced in the original interaction is minimal. Also, we choose to limit modifications of off-the-shelf games to a single continuous time rate variable and use solely the design of the controller to optimise for rehabilitation goals.

From a computational perspective, the interaction is represented on Figure 4.3. In this particular instance, we chose to parametrise our control modality with a single parameter that represent the distance (SPREAD) between the control regions. By changing the value of SPREAD, the motions required by the players are impacted: an increase in SPREAD will increase the range of motions or reaching. To adapt for a lower control afforded to the player by this new control input, the time rate (T_RATE) of the game is used as an optimisation variable. This is a simple yet robust game design pattern for single player games, commonly used to adapt for difficulty.

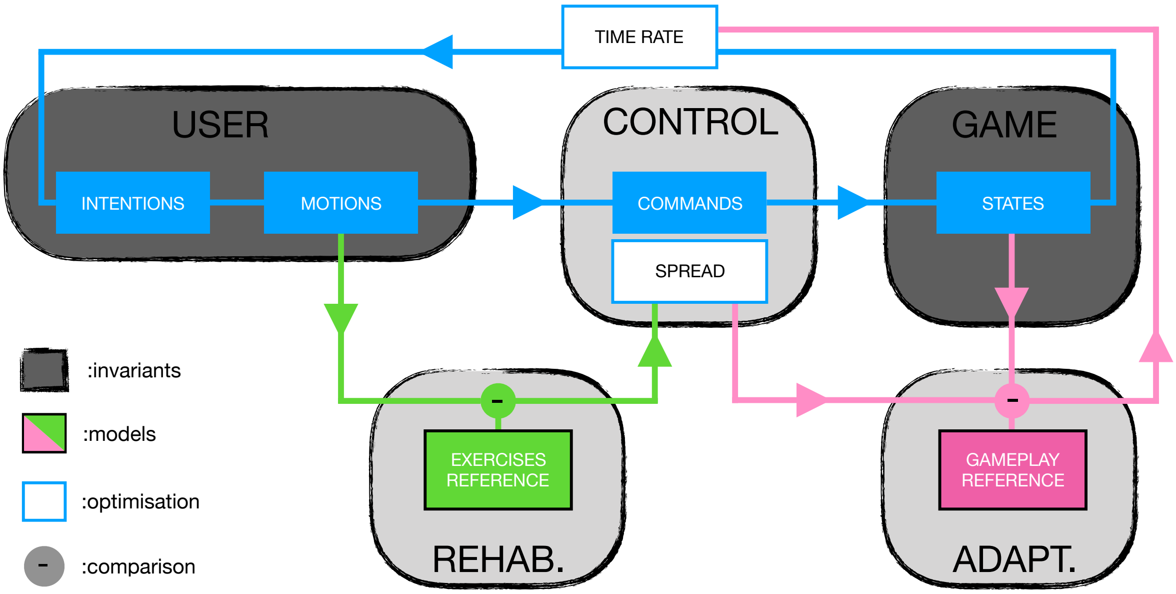

Figure 4.3: Simplified block diagram for the interaction. In blue, the loop that generates engagement, in green the loop that satisfies the rehabilitation goals, in red the loop that adapts the level of difficulty against a model of reference gameplay.

4.4 Research Plan

The design of a control modality will impact the overall user performance. In this instance, we are interested in upper limb motions while games are usually played with a fingers based controller. Figure 2.2 has shown that in a pointing task, these two limbs afford very different performance capabilities in term of information throughput. It is safe to assume that the trend will hold for the task of pushing controller buttons. We can also assume that Arcade games take players close to their control limits by requiring precise and frequent input. In other word, where players are bounded by their control ability, a reduction in this ability will have an impact on performance. The requirement of adapting exercises could also have an impact on performance. For example, if we model the task of pushing controller buttons as successive pointing tasks, eliciting further reaching motions to activate the controller buttons, or increasing D will increase the difficulty of the task, according to Fitts' law. Here, we have proposed to use the time flow of the game to afford users more time to perform the actions required by the gameplay. It is however unknown what level of reduction would achieve this objective.

The first research question is to find out what is the effect of new input control on experience and rehabilitation goals and what is the effect of time rate on experience and difficulty.

- **RQ1.0**: Can the control of time rate counterbalance the effect of a new input modality and the effect of different exercises levels?

- **RQ1.1**: Can the design of the control modality manipulate the rehabilitation intensity and impact the behaviour of users?

The question on how to tune the design will be addressed in the second part of the chapter.

4.5 Experiment with Unimpaired Participants

We designed an experiment to investigate research questions **RQ1.0** and **RQ1.1**.

These questions can be answered without loss of generality with unimpaired participants. Despite the differences between the targeted users who have sustained a spinal cord injury and potential participants of the user study, the main effects should be observable with both groups. The difference in throughput between arm and finger control, as well as the influence of control spacing on performance might only have different magnitude between groups.

4.5.1 Apparatus

We used an Optitrack system for the tracking of the participants hand. The interaction surface was afforded by an office desk table with a width of $78cm$. The output from the Optitrack system was redirected to another computer running the game of *Pac-Man* in a browser. We used an opensource version of the game that was re-created with authenticity in mind[3]. The game was instrumented to expose some of its internal state variables: the commands received and performed by the game, the score, the number of the avatar steps taken in multiples of 10 and the avatar direction changes were logged as individual events along with their frame number and time stamp. These measures of game state, also not automatically exposed by the game, are in general easily accessible.

4.5.2 Task

To limit the experiment's length to under an hour, we used the first level of the game *Pac-Man* as the task. The task was completed when all the pellets were consumed in the level or when all 3 lives had been used. It was thus not possible to fail the task.

4.5.3 Design

RQ1.0 and **RQ1.1** can be answered with a factorial design that measures potential effects and interactions of the independent variables gamepad control position (SPREAD) and time

[3]see `https://github.com/masonicGIT/pacman` for the source code and a discussion about the accuracy of this remake.

flow (T_RATE) on each participant task end-score (SCORE). SPREAD was evaluated with 2 levels ($10cm$ and $40cm$) covering the expected range of motion in rehabilitation and T_RATE was evaluated with 3 levels ($T_r/3$, $2T_r/3$ and T_r), see Table 4.1. In practise, setting T_RATE was achieved by altering the number of game updates per seconds. These values were chosen to enable a comparison with the game's original time rate and to lower the overall information throughput requirement for lower values since we expected the challenge to be increased when participants used their arm instead of their fingers. As the performance of individual participants was likely to exhibit a high variance, we opted for a within subject design. To counterbalance potential learning effects in playing the game and adapting to our novel interaction paradigm, we used a balanced Latin square design. With two independent variables (IV) SPREAD and T_RATE evaluated with respectively two and three levels, we obtained six testing conditions.

IV	levels
T_RATE	$T_r/3$, $2T_r/3$, T_r
SPREAD	$10cm$, $40cm$

Table 4.1: Values for the independent variables T_RATE and SPREAD. T_r is the default time rate of the original game.

Finally, to allow for the measure of a gameplay reference, a pre-test and a post-test were included to the experiment. Participants were tested with a conditions using a keyboard (KEYBOARD), by opposition to conditions using our system (TRACKER). The pre-test and post-test included two repetitions. Each block was designed with three repetitions provided the third repetition (e.g. game) was started under the 8 minute mark in that block. The design of the experiment included thus: 12 participants $\times$ (2 SPREAD $\times$ 3 T_RATE $\times$ 1 BLOCK $\times$ [2-3] Repetitions $+$ 2 PRE-TEST + 2 POST-TEST), equivalent to 192 to 264 total trials.

4.5.4 Procedure

Participants were welcomed and provided with the ethics and information sheet. Upon acceptance to participate, they received a £5 compensation. Participants were asked for demographic information and about their previous experience with *Pac-Man*. They were given a short explanation about the game mechanics, such as the need for consuming all the pellets to complete a level, the role of the ghosts and the effect of the power-ups placed at the corners of the level. They were seated in front of a table, introduced to the tracking system, and equipped with a pair of headphones to receive the system's audio feedback. An optical marker was attached with a stretchable band adaptable to different users' hand physical features. They were asked to acquire each of the gamepad controls once. The centre of the gamepad was chosen as a full forearm extension from the table edge adjacent to their trunk and, laterally, equidistant from both table edges. It was indicated on the tabletop with a protruding marker that was taped in place for each participant. The participants were encouraged to produce their best possible score. To produce incentive for performance, we added an extra prize of £15 awarded to the participant that would produce the highest score in any of the testing conditions. At the end of each condition, participants were offered a break. Participants were responsible for advancing through the experiment by starting a new game. They were informed when they were changing testing condition but were not told about the current value of SPREAD or T_RATE. At the end of the experiment, participants were invited to provide qualitative feedback during a 5 minutes discussion.

4.5.5 Participants

12 participants served in the experiment: 4 females and 8 males, 1 of whom was left-handed, with a mean age of 29 ± 7. The study was approved by the University of X ethics committee. No participants presented any mental or physical disabilities.

4.5.6 Results

Data Processing

For each testing conditions, we recorded the logs from the game of *Pac-Man*. From these, end-task SCORE was computed as the highest score at trial completion.

Learning effects

Participants reported their experience with playing the game of *Pac-Man*. All participants reported having played before the experiment, even if some reported that they could not remember when was the last time. We looked for some learning effect between the PRE-TEST and POST-TEST conditions on KEYBOARD. A statistical analysis showed no significant effect of game number on SCORE. However, we observed an increasing trend in SCORE with a mean value of 3697 points for the first PRE-TEST game and a mean value of 5077 points for the last POST-TEST game. The mean value over all games was 4390 points with a standard variation of 950 points. The maximum score was obtained in the last POST-TEST level with a value of 8200 points. Refer to Table 4.2 for numerical values.

Normalisation

We computed the values for SCORE per participants. The distribution of SCORE across participants was diverse. Some participants, such as (0, 7 and 10), produced a very low

	PRE		POST		
game #	1	2	3	4	all
mean	3697	3992	4795	5077	4390
std	978	1438	1529	1705	950

Table 4.2: Mean value and standard deviation for SCORE over PRE-TEST and POST-TEST levels and averaged over games (all) in the last column.

variance while others, such as 5, presented a much bigger range of SCORE. We computed a normalised value for SCORE, marked as NSCORE in the following, by dividing the SCORE value with the average value of 4390 points obtained over the PRE-TEST and POST-TEST levels, NSCORE=SCORE$/mean($SCORE$)$. The results across participants are presented on Figure 4.4, with the values for the mean and standard deviation of SCORE, NSCORE in Table 4.3.

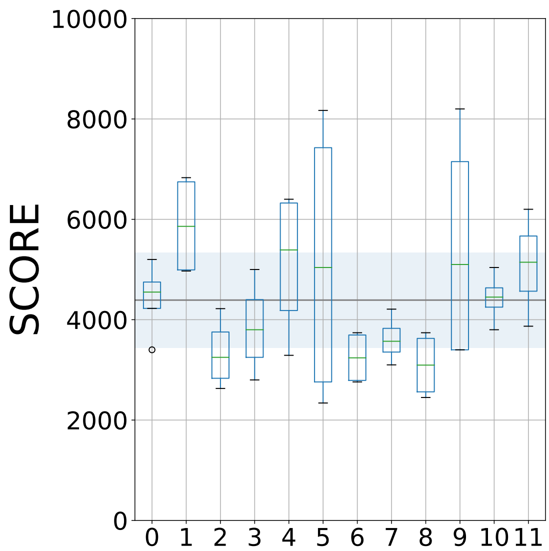

Figure 4.4: Distributions of SCORE across the 12 participants in KEYBOARD.

	SCORE	NSCORE
mean	4390.0	1.0
std	950.0	0.22

Table 4.3: Mean and standard deviation for SCORE and NSCORE.

Outcome Effects

The results for NSCORE as function of T_RATE and SPREAD are plotted in Figure 4.5, while the data containing the mean and standard deviation is reported in Table 4.4 for NSCORE.

A repeated measure two-way ANOVA on NSCORE with SPREAD and T_RATE as factors showed a significant main effect of T_RATE ($F_{1.47,16.15} = 22.11$, $ges = 0.41$, $p < 0.0001$),

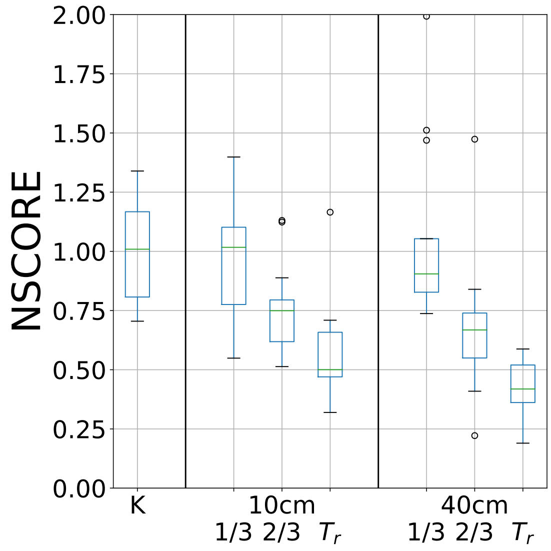

Figure 4.5: Effect of SPREAD and T_RATE on NSCORE for KEYBOARD (K) and TRACKER. The graph includes three parts, with KEYBOARD (K) on the leftmost subplot, and TRACKER with the six combinations (Table 4.1) grouped by value of SPREAD on the middle and rightmost subplots. Values for SPREAD and T_RATE are reproduced on first row and second row, respectively.

NSCORE	$T_r/3$	$2T_r/3$	T_r
$10cm$	0.96 ± 0.24	0.75 ± 0.21	0.57 ± 0.22
$40cm$	1.05 ± 0.39	0.68 ± 0.30	0.42 ± 0.12

Table 4.4: Mean and standard deviation of NSCORE as a function of SPREAD and T_RATE in rows and columns, respectively.

no significant main effect of SPREAD and no interaction T_RATE $\times$ SPREAD on NSCORE. NSCORE was negatively correlated with T_RATE, decreasing from, on average, 96% of the reference keyboard SCORE at $T_r/3$ to 57% at T_r. Although not significant, SPREAD appeared to correlate negatively with NSCORE for T_RATE levels of $2T_r/3$ and T_r, but not for T_RATE level of $T_r/3$. We ran pairwise comparisons adjusted with Holm-Bonferroni on all levels with the inclusion of KEYBOARD. It showed no differences between KEYBOARD and $T_r/3$. Supporting the lack of significant effect of SPREAD on NSCORE, we found no differences between any SPREAD values for the same T_RATE. We also found no differences between levels $(T_r, 10cm)$, $(2T_r/3, 10cm)$ and $(2T_r/3, 40cm)$. In other words, participants performed as well using the tracker as with a keyboard, measured by SCORE, provided T_RATE was reduced accordingly. The range of T_RATE values that produced an equivalent score is however unknown, but most likely located around a value of T_RATE close to $T_r/3$.

Effect on Hand Positions

The hand positions for the lowest and highest performing participants in their worst, median and best games are plotted on Figure 4.6. The testing conditions under which these were produced are written in brackets over the subplots as [T_RATE SPREAD]. These traces reveal different pointing strategies between both participants. The participant on the top row always passed back through the centre in between motions towards the actionable areas, while the participant on the bottom row made direct connections between actionable areas.

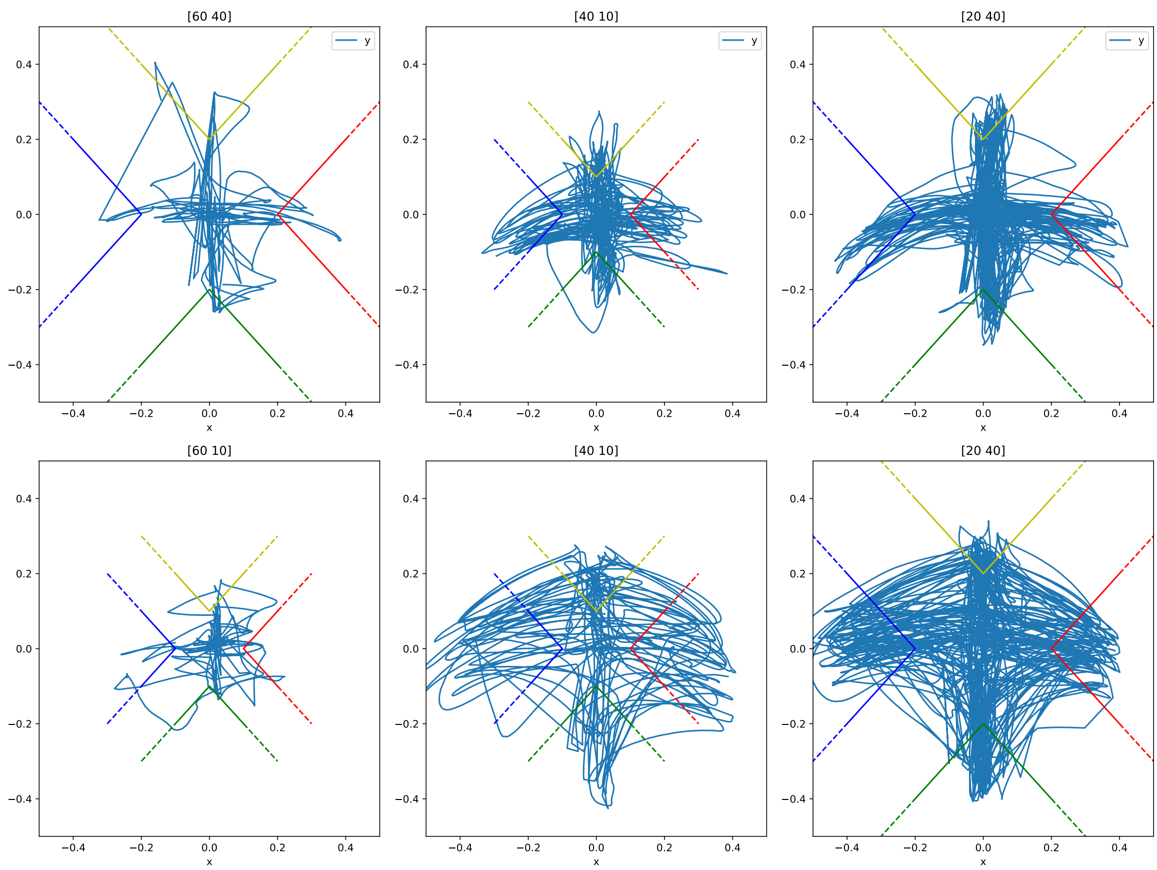

Figure 4.6: Traces of hand positions. Lowest and highest performing participants on the top and bottom row, respectively, with their worst, average and best games on column 1, 2 and 3, respectively. The inner contour of the actionable controls are equivalent to those on Figure 4.2.

The hand distance for the rest position was also computed per conditions. It is plotted, grouped by SPREAD, as an histogram on Figure 4.7. The distribution was bimodal, displaying a resting position (leftmost mode) and a targeting position (rightmost mode). It was fitted with a Gaussian mixture model [57] with two components. The mean and covariance of the Gaussian distributions representing the targeting positions are reported in Table 4.5. There was an increase in the reach from participants with a targeting position increasing from

$20.3cm$ on average to $25.4cm$ on average when SPREAD was increased from $10cm$ to $40cm$. The increase in reach was not proportional to the increase in SPREAD, which showed that participants were overshooting the boundary of the actionable areas for the smaller value of SPREAD.

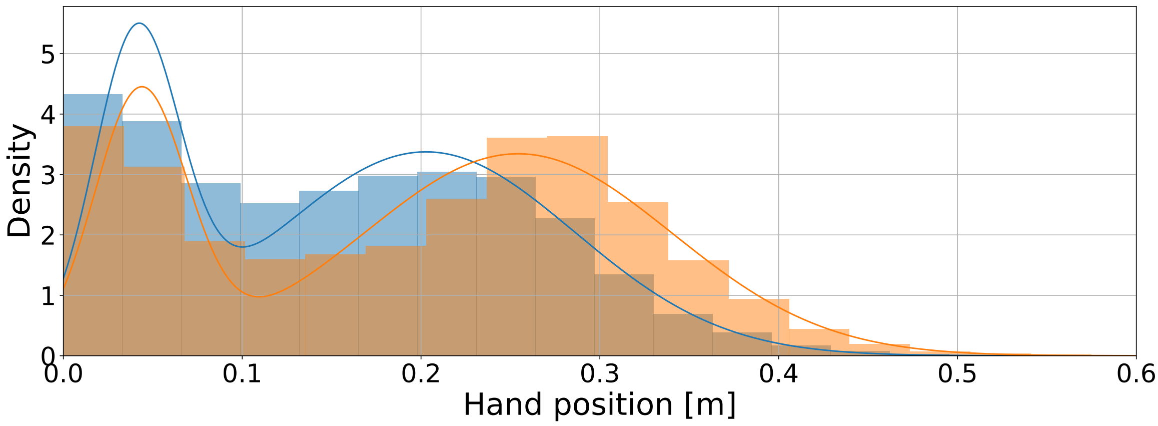

Figure 4.7: Distribution of the distance of participant hand positions from their mean position for SPREAD equal to $10cm$ and $40cm$, in blue and orange, respectively.

2nd Gaussian	$10cm$	$40cm$
mean	20.3	25.4
std	8.3	8.6

Table 4.5: Mean and standard deviation in centimetres for the Gaussian distributions representing the targeting positions for both values of SPREAD $10cm$ and $40cm$ in blue and orange, respectively.

4.5.7 Qualitative results

From the Participants

A short discussion followed the experiment where participants were asked “how did they feel about the different conditions”. The feedback was collected on a notebook as participants were expressing themselves, with the most salient quotes selected by the interviewer. Some preliminary insights can be extracted from this fragmented qualitative feedback.

The mention of workout in participant feedback seems to indicate that the interaction was having an impact on exertion. Regarding the level of physical involvement required by the interaction, three participants mentioned that the experiment was a “workout”, one of them being more specific and declaring being “a bit exhausted in the shoulder”. Even though all participants were offered breaks during conditions, only one of them did effectively mark a pause in the experiment for five minutes.

Some feedback regarding specifics on the design were that an asymmetry in terms of controls position was expressed by three participants who voiced that it was difficult to reach the virtual control at the top of the table, which activated the UP command, saying that it was either “too high” or “quite difficult”. Regarding the control that was placed too high to be easily actionable, this could call for a better placement to take into account the fact that it is harder to reach further up than left and right. Also, one participant mentioned that “sliding across the reference felt weird.” This refers to the protruding marker indicating the centre of the controls. Specific skills were pointed out by two participants, one of whom acknowledging “not being quick enough to make two turns” while the other declared she “can’t turn off a corner”. One participant did mention not realising that the “change in position” across testing conditions.

Some participants became involved in the experience proposed by the experiment, and were “into the game”, had “fun” or realised that the experiment “did not feel like an hour”. However, the level of enjoyment did fluctuate for some participants across the different levels. Some difference in the gameplay proposed was also pointed out by participants with a particular mention on T_RATE level $T_r/3$ were three participants described the game as “not exciting”, “a bit too easy” or presenting “not much challenge”, while the T_RATE level T_r was described as “horrible” by one participant.

From the Occupational Therapists

The OTs have been involved in the design process and with some testing of the system. The first qualitative feedback was positive with regards to the motions that were elicited from gameplay session with the system and the perceived level of enjoyment as compared to their non-digital exercises.

4.5.8 Conclusion

This first experiment was meant to answer the following research questions:

- **RQ1.0**: Can the control of time rate counterbalance the effect of a new input modality and the effect of different exercises levels?
- **RQ1.1**: Can the design of the control modality manipulate the rehabilitation intensity and impact the behaviour of users?

Both questions have been answered positively. The variable time rate was shown to allow for setting the game difficulty at a level that was producing the same score as an interaction with a keyboard. It could thus be assumed that the user engagement would be maintained with that setting. A short qualitative survey has revealed that the interaction was enjoyable, even though it was described as tiring. The position of the actionable control also had a positive effect on the user motions with an increase in reach when the controls were set apart. The OTs have also had the opportunity to assess the similarity between the motions produced and the ones they were expected and were satisfied with the results.

4.6 Design Optimisation

The previous experiment answered the first research questions positively. However, the practical use of the system in-situ remains challenging. To be precise, patients having sustained a spinal injury will likely exhibit varying performance across participants. The relationship recorded during the previous experiment between T_RATE and SCORE should be similar but the value of $T_r/3$ might not be adapted, as these were recorded with unimpaired users.

To illustrate the challenge, a hypothetical scenario can be used.

Alex is scheduled to go through a first rehabilitation session with his occupational therapist Alice. After a quick assessment of his motor capabilities, Alice proposes to work on Alex's reach capability of the upper arm and engages in a session that revolves around playing the game of Pac-Man. Some virtual controls are set on the tabletop, in lieu of the position that Alex's hands are meant to reach. The game itself is hard, and Alex's reach capabilities, even though they have been assessed during the first minutes of the meeting, are not known down to his capacity to perform fast motions consistently. A few minutes after the game is started, Alex starts to tire and keeps losing in game after only a few seconds of play. Following some complaints to Alice, the speed of the game is reduced to a fraction of its value. However, the game is now too easy for Alex: the challenge is not present any longer and he quickly loses interest in the task.

This hypothetical scenario is meant to point out that the tuning the game difficulty in a timely manner, beyond setting it properly in the first place can be a challenge. A naive but valid approach to this problem would be to use the game score. After all, Arcade games, and *Pac-Man* is no exception, do rely heavily on score as a principal feedback mechanism.

The score structure in *Pac-Man* is organised with points awarded each time the player consume an object by moving over it. 240 pellets grant 10 point each, bonuses in the form of fruits are worth anywhere between 100 and 5000 points and ghosts, which can also be consumed when the player is in a power-up phase, award each between 200 and 1600 points. The power-up phase is accessed by consuming one of 4 special pellets. The ghosts point structure is exponential: the first one grants 200 points, the second one 400 points, then 800 points and finally 1600 points. The issue with this point structure is that the score of a given player heavily depends on its strategy and capacity to consume ghosts. The rate at which point is accrued is also very variable over time, due to the presence unlikely events with big point returns. As a result, using score to set the game difficulty might be problematic since a complex model of the patient's performance in *Pac-Man* becomes necessary.

The goal for the computational design, see 4.3.1, is to quickly find configurations that minimise the cost function. Here the cost function would be the difference between the keyboard score and the experiment score. A suitable configuration is one that minimises the difference between both scores. To assess this difference, since both variables are distributions and exhibit some variance, a number of samples is required before a decision can be made. This is a similar idea to the statistical analyses that were carried out earlier in this chapter. However, the value for score is available only once at the end of a game. This limits the usefulness of score as the basis for defining the cost function, due to its variation across individuals and its sporadic availability.

Another avenue for this problem comes to mind by asking the following question: where does motivation in games come from? Przybylski et al. [101] have looked at a motivational model of video game engagement and distinguished two types of motivation: intrinsic motivation where behaviours are pursued for their own sake and extrinsic motivation where behaviours are pursued to access desired end states or to avoid aversive ones. Focusing on the former, as accepted in the literature, they point out that "activities foster greater intrinsic motivation to the extent to which they satisfy three fundamental human needs: the need for competence (sense of efficacy), autonomy (volition and personal agency), and relatedness (social connectedness)." Arcade games in that space are excelling at supporting the need for competence and address the critical need of balancing of game difficulty and player skill. Also, Przybylski et al. elaborate that the "mastery of controls plays an important role in game motivation, largely as a necessary, but not sufficient, condition for achieving psychologically need-satisfying play." In a related attempt to understand enjoyment in games, the role of the game controller itself has been investigated. Using the flow framework [102], Limperos et al. [103] compared the difference a Wii and a Playstation controller produced on the game experience, with the Wii controller being qualified as "natural" and requiring more involvement from the user body. They showed that the performance as defined as the end-game score was not alone explaining the enjoyment, and that the sense of control that users experienced was rather a much more salient indicator. In their concluding remark, the sense of control and the need for competence were linked together.

The observations from the hypothetical scenario and the findings from these studies point to another direction. It appears that the sense of control would provide a link to enjoyment[4] but

[4] *GameFlow* [104] is a model for evaluating player enjoyment in games but it does not provide a real-time access to measurable quantities, and is conceived as an evaluation tool for assessing games after a thorough analysis.

a measure for it needs to be crafted. As a result, a new research question can be investigated:

- **RQ1.2**: Does a measure of user behaviour inspired by the sense of control allows for low-latency sampling and user independence?

4.7 Modelling User Behaviour

We have established that a measure of control could be useful for helping practitioners to set the game speed via T_RATE to an adequate level. That measure should also abstracts the potential differences between users, in terms of strategy for example, and focus on their capability to play.

We decided to take inspiration from the notion of empowerment, stemming from [105]. In this view, users behaviour is about controlling their perceptions: the more control they have over their perceptions, the more empowered they are by the user interface to achieve their goals. This approach has shown to be successful in HCI, for example in a tracking task [106] where it correlated with performance. Here, we base our measure of control on a simple interpretation of the empowerment idea: linking user action to game outcome. We are seeking to build a reference model of play based on distributions of variables that relate to user actions and user effects which in turn give access to a measure of similarity between a reference and the observed behaviour from an unknown player.

For the model of user's action, research on text-input shows that the number of key presses in a time interval has been successfully modelled as a Poisson distribution, and its continuous counterpart defined through the time difference between two consecutive presses or inter-key interval has been shown to follow an exponential distribution. The assumptions being that events occur continuously and independently at a constant average rate λ. We do not know whether the key presses needed for enacting the gameplay will follow the same distribution, but due to the similarity of the input modality such model will be investigated.

For the user's effect, the measure chosen depends on the game being played. After observations of game sessions of *Pac-Man*, we noticed that players with poor avatar control were more likely to get stuck in walls with their avatar staying motionless for small period of time. To encode this realisation, we decided to measure the *Pac-Man* turning time, or the time taken by the avatar to perform a direction change. If the user's effect is immediate, only 1 frame is necessary. If the effect is less effective, more frames will be needed. For that model, the distribution which models the best the data is unknown and will be informed by empirical evidence.

In summary, the inter-key interval (IKI) is the difference measured in frames between two keypress events. It encodes the user input rate capability and is a measure regularly used in research related to typing. The *Pac-Man* turning time (PTT) is the time interval measured in frames between direction changes of the avatar. This variable is more specific to the chosen game, and relates to the players' capability to avoid getting stuck at walls.

Keyboard reference model

One of the strengths of this model is its simplicity. The data needed to build it can be collected from normal play sessions. The data used here is taken from the the PRE-TEST and POST-TEST conditions on KEYBOARD of the previous experiment, see section 4.5. This allows us to present the real values for our fitted models.

We decided to consider both variables PTT and IKI as continuous random variables, even though they take discretised positive values. The shape of the distribution for PTT resembled an Exponential distribution [57], with only positive values, a sharp peak around zero and a decaying trend towards higher values. The probability density function (pdf) is given by:

$$Expon(x|\lambda) = \lambda \exp(-\lambda x)$$

and the parameter ($\lambda = 2.2$) gave the best fit.

The distribution of IKI seemed to be well modelled with a Laplace distribution [57]. This distribution describes the difference between two independent identically distributed exponential random variables, hinting that key presses were following an Exponential distribution similar to PTT. The pdf is given by:

$$Lap(x|\mu, b) = \frac{1}{2b} \exp(-\frac{|x-\mu|}{b})$$

and the parameters estimated from the data were ($\mu = 26.0$, $b = 17.3$).

The empirical distributions and fitted models are shown in Figure 4.8 with the data associated in Table 4.8. Histograms (in blue) represent the empirical distributions, with their associated pdf shown in black. We can see that the models are approximating well the empirical distributions.

RV	*distribution*	*parameters*
PTT	Exponential	$\lambda = 2.2$
IKI	Laplace	$\mu = 26.0, b = 17.3$

Table 4.6: Distribution types and estimated parameters for the variables PTT and IKI.

Our reference model $\theta_{ref} : (\mu_{IKI}, b_{IKI}, \lambda_{PTT})$ is thus defined by the parameters of both distributions. The association of both variables is designed to limit the model's reliance on

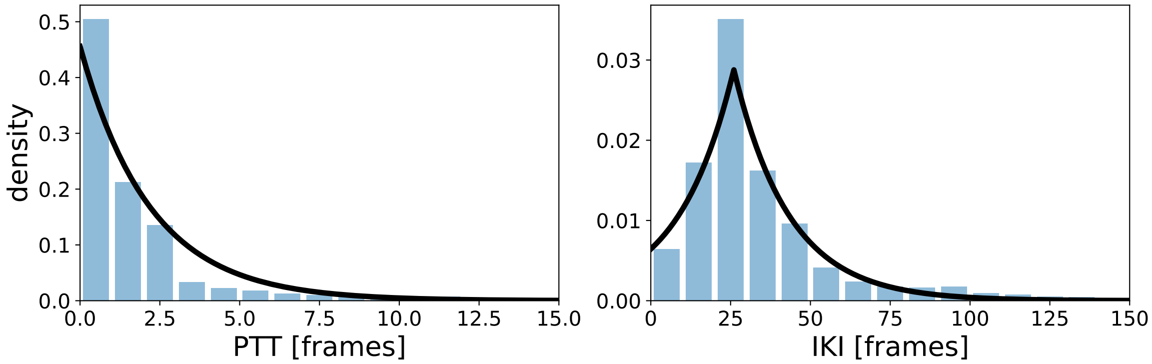

Figure 4.8: Empirical (blue histogram) and fitted (black pdf) distributions for variables PTT and IKI.

user specific behaviour, such as tactics or strategy and focus on the input capability and its effect on the game state.

Given the observed play session of a user, this model can be used to estimate the proximity between the user behaviour and our reference keyboard behaviour. We gather the observations (IKI, PTT) over F frames into $\mathbb{X} : (x_i^{IKI}, x_j^{PTT})$ of lengths (M, N). We use the likelihood $\mathcal{L}(\theta_{ref}|\mathbb{X})$ of our reference model given the observed data as a measure for inferring how likely a user is to be behaving under the assumptions of our reference model:

$$\mathcal{L}\left(\theta_{ref}|\mathbb{X}\right) = p_\theta\left(x\right)$$

Assuming independence between X_{IKI} and X_{PTT}), we have:

$$p_\theta(x) = p_{X_{IKI}}(x) \times p_{X_{PTT}}(x)$$

and computing the per-sample likelihood, for an observation of N sample yields:

$$p_\theta(x) = \sum_i^N p_\theta\left(x_i\right)/N$$

Finally, the per-sample log-likelihood of our model given $\mathbb{X}$ is then defined as:

$$\begin{aligned}\log \mathcal{L}(\theta_{ref}|\mathbb{X}) = &\sum_i^M \log\left(p\left(x_i^{IKI}|\mu_{IKI}, b_{IKI}\right)\right)/M \\ &+ \sum_j^N \log\left(p\left(x_j^{PTT}|\lambda_{PTT}\right)\right)/N\end{aligned} \tag{4.4}$$

where $p(x_i^{IKI}|\mu_{IKI}, b_{IKI})$ and $p(x_j^{PTT}|\lambda_{PTT})$ are the densities computed on the reference model. This quantity is a measure of how likely the reference model is of explaining the data in a given observation and can be used for comparisons between different observations.

With a behavioural model for users, we can now form two hypotheses:
Ha: This measure should be more user agnostic than a measure based on the in-game score.
Hb: This measure should constitute a proxy for the level of user control.

4.7.1 Normalisation

The values for the log likelihood (LL) were computed according to equation 4.4. In the same fashion as SCORE, one value per game was obtained. We also computed a normalised value for LL, marked as NLL in the following, by taking the inverse of LL and multiplying it with its average value obtained over the PRE-TEST and POST-TEST levels, NLL$= mean(\text{LL})/\text{LL}$, similar to how NSCORE was computed. Figure 4.9 shows the distribution of NSCORE per participants across participants. It should be compared with Figure 4.4.

We observed a clear difference in standard deviation between NSCORE and NLL with a value of 0.22 and 0.06, respectively, see Table 4.7. For NSCORE, we have five participants (1, 2, 4, 6 and 8) whose average NSCORE lies father than one standard deviation from the overall mean, while for NLL only two participants (6 and 7) present the same deviation from the mean. In other words, our synthetic model of behaviour is less noisy than using the score as a measure of performance.

This result validates hypothesis **Ha** about the independence of NLL in terms of users. What need to be investigated now is whether NLL do vary with T_RATE and SPREAD.

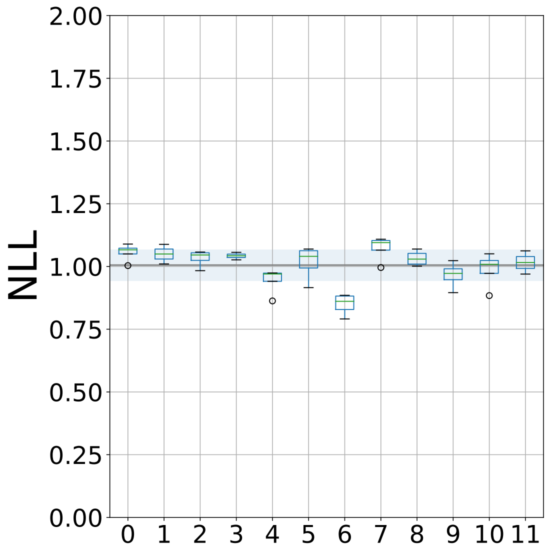

Figure 4.9: Distributions of NLL for all participants over the PRE-TEST and POST-TEST conditions.

	NSCORE	NLL
mean	1.0	1.0
std	0.22	0.06

Table 4.7: Standard deviation for NSCORE and NLL.

4.7.2 Outcome Effects

The effect of SPREAD and T_RATE on NLL was also investigated, see Figure 4.10. It should also be compared to the effect of SPREAD and T_RATE on NSCORE, see Figure 4.5. The figure contains six conditions and the data containing the mean and standard deviation is reported in Table 4.8 for NLL.

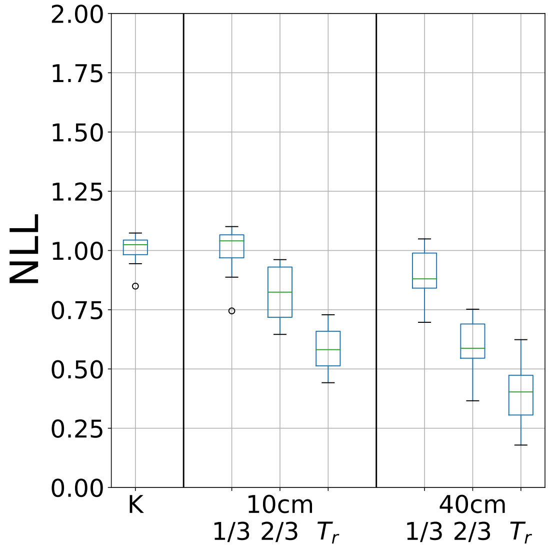

Figure 4.10: Effects of SPREAD and T_RATE on NLL for KEYBOARD (K) and TRACKER. The graph includes three parts, with KEYBOARD (K) on the leftmost subplot, and TRACKER with the six combinations (Table 4.1) grouped by value of SPREAD on the middle and right-most subplots. Values for SPREAD and T_RATE are reproduced on first row and second row, respectively.

NLL	$T_r/3$	$2T_r/3$	T_r
$10cm$	1.00 ± 0.10	0.81 ± 0.12	0.58 ± 0.10
$40cm$	0.89 ± 0.11	0.60 ± 0.11	0.40 ± 0.12

Table 4.8: Mean and standard deviation of NLL as a function of SPREAD and T_RATE on rows and columns, respectively.

Similarly to NSCORE, we ran a repeated measure two-way ANOVA on NLL with SPREAD and T_RATE as factors. The analysis showed a significant main effect of T_RATE ($F_{1.96,21.58} = 83.34$, $ges = 0.76$, $p < 0.0001$). The analysis also showed a significant main effect of SPREAD ($F_{1,11} = 72.50$, $ges = 0.39$, $p < 0.0001$). We observed a negative impact of T_RATE and SPREAD on NLL with lower values with increasing values of T_RATE and

SPREAD. We ran pairwise comparisons adjusted with Holm-Bonferroni on all levels with the inclusion of KEYBOARD. We found that only level ($T_r/3$, $10cm$) was similar to KEYBOARD. We also found some similarities between pairs ($T_r/3$, $10cm$) and ($T_r/3$, $40cm$) and pairs ($T_r/3$, $40cm$) and ($2T_r/3$, $10cm$) showing that similar effect on NLL can be achieved with different parameters pairs. Finally, the analysis found a small interaction T_RATE $\times$ SPREAD ($F_{1.58,19.39} = 5.28$, $ges = 0.04$, $p = 0.02$). The reason seems to be the relatively high value of NLL for level (T_r, $40cm$). However, the amount of data collected in this condition is comparatively smaller than other conditions: the poor user performance had the result of shortened game sessions. This effect is thus likely an artefact.

From these results, it is important to note that for condition ($10cm$, $T_r/3$), the effect of T_RATE and SPREAD on NLL allowed to match the behaviour of the KEYBOARD condition. This is similar to what was observed in section 4.5.6. In other words, a completely different model than one based on SCORE, relying instead on low-level variables of user behaviour, did find the same solution for the computational optimisation problem. One difference between both models is found with the condition ($40cm$, $T_r/3$) which is found to be different for NLL but not for SCORE. The fact that NLL presents a much lower variability than SCORE could explain this difference. However, it appear thus that there is a link between both models.

Correlation of NLL with NSCORE and Sampling Period

Given the similarities between the effect of SPREAD and T_RATE on both NSCORE and NLL, it is opportune to check for a possible correlation between the two variables. A scatter plot of their associated values for all measures except for the keyboard condition is plotted on Figure 4.11 on the left. A linear regression reveals a significant relationship with ($slope = 0.83$, $intercept = 0.14$, $stderr = 0.07$) and a R^2 value of 0.39.

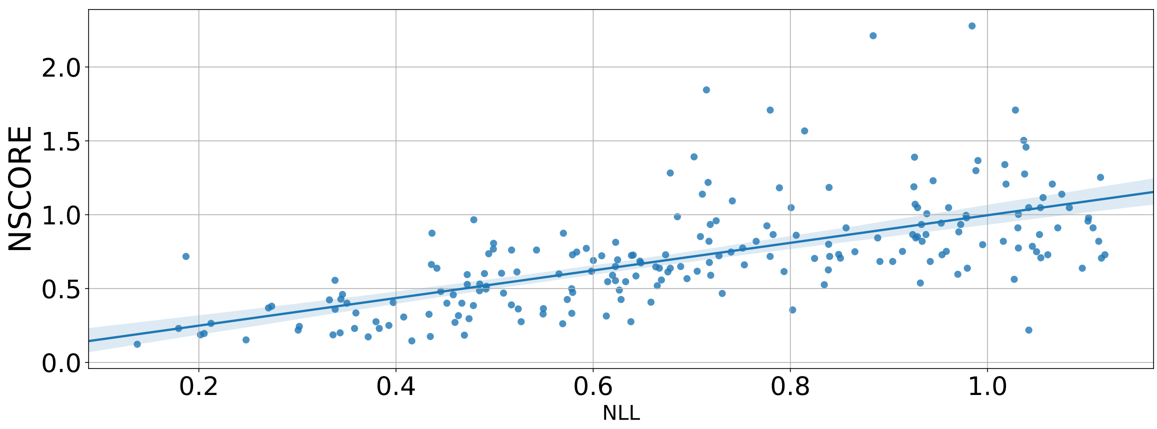

Figure 4.11: Correlation between NLL and NSCORE.

The relationship between both variables shows that it would be possible to use one for the other in an optimisation scenario. This result validates the hypothesis **Hb** which was advancing a link between our proposed model and the user sense of control. Indeed, it shows that if the behaviour of participants, in terms of their behaviour as measured by IKI and PTT, is maintained to a level equivalent to what was observed on a keyboard, then the score should also be maintained. The assumption that maintaining the score to a level similar to that of keyboard reference gameplay is central to this chapter, as score is used as the proxy for engagement.

Finally, one of the reason for designing such model was the availability of many more samples than values for score. We measured from the TRACKER condition the time elapsed between samples for SCORE and LL, see Figure 4.12 and Table 4.12. Obtaining one sample for SCORE took on average 2210 frames, while one sample for LL was available every 40 frames on average. This answers the low-latency part of the last research question **RQ1.2**. The measure of user behaviour based on low-level variable of gameplay provides 55 times more samples per unit of time than the counterpart model based on SCORE.

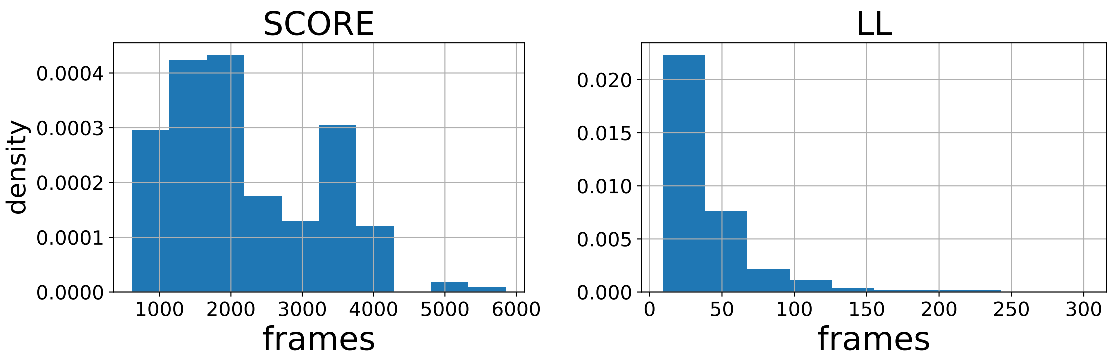

Figure 4.12: Distribution of time taken to measure on sample from SCORE and LL on the left and right, respectively. LL

	SCORE	LL
sampling period [frames]	2210 ± 1057	40 ± 31

Table 4.9: Average sampling period in number of frames (with standard deviation) for SCORE and LL. This represents the waiting time before a new sample is available, which is roughly 55 times longer for SCORE than for LL.

4.7.3 Conclusion

The experiment with unimpaired participants has allowed us to establish that a new control modality can be used without an impact on the performance as measured by the game score provided the framerate of the game is lowered. We also verified that there is an increase in the arm reach of participants when the distance between the actionable areas is increased. Finally, we established that a specifically designed measure of performance was proven to exhibit low-latency and user independence. Reinforcing the idea that it links to the user sense of control, it was also shown to correlate well with the user in-game score.

4.8 Preliminary Experimentations In-situ

In parallel to the laboratory experiment, some preliminary tests with users with SCI were conducted. We also developed a portable system for testing purposes in the ward of the QEU hospital.

Moving from a laboratory scenario to a real-world scenario usually presents a new set of challenges, leading to some adaptations. The system is based on optical tracking, which was relying for the previous user study on using an Optitrack system. Even though Optitrack is state of the art with regards to precision in tracking, it is very expensive as well as bulky and hard to setup. As a result, the system was built around a depth camera providing the tracking of a reflective marker which was placed in the same manner on the dominant hand of the player. A set of workshop has been conducted to ensure that the OTs were able to install the system by themselves. The OTs expressed the need for exercises with very small range of motions, which corresponded to users at the beginning of their rehabilitation. The smallest condition in the previous experiment ($10cm$) was deemed to big. The system was thus fitted with a graphical interface for an interactive positioning of the virtual gamepad with its parameters as well as tuning the tracking parameters. In this version, we enabled the parameter SPREAD to be controlled along the two principal axis of the gamepad, from $0cm$ and above. The feedback from the occupational therapists was again positive with regards to the motions that were elicited from gameplay session with the system. Ethical approval has been obtained but no further testing has been conducted unfortunately as of the writing of these words.

Two users with SCI were approached during a workshop organised by the MoreGrap project. One participant had suffered from a very high level SCI and did not have remaining motor functions in the trunk. Given the lack of upper body stability, reaching over a tabletop to interact with the virtual control presented a problem. The portability of the system enabled the placement of the control as close as possible to the participant. The wheelchair was positioned sideways so that the participant's elbow almost touched the tabletop. The control area was also tilted to accommodate for the new principal interaction axis. This interaction showed that every patient might require specific configurations, and that the portability of the system was important. The other participant had been engaged in rehabilitation for more than a year after the injury and exhibited a relatively good control of his forearm. He was capable of moving his hands to the specific locations of the virtual controls with precision and speed. This participant mentioned that he could see a parallel between the physical exercises he had to perform during his rehabilitation and the ones that were elicited by the gameplay. Note that this participant was aware that this system had been designed with rehabilitation goals in mind. Nevertheless, this comment is consistent with the feedback from the OTs, see 4.5.7.

In summary, both participants were capable of interacting with the system. They reported having enjoyed the experience and saw some connections to their past experiences with upper limb rehabilitation.

4.9 Conclusion

We had defined three research questions:

- **RQ1.0**: Can the user performance be maintained across different control modality (keyboard and optical gamepad) and for different value of spread as measured by the game score?
- **RQ1.1**: How do users adapt their behaviour when gamepad controls are spread further apart, and does it increase their reach or range of motions ?
- **RQ1.2**: Which measure of performance allows for low-latency sampling and user independence?

And for the synthetic measure of performance we also had emitted some hypotheses:
Ha: This measure should be more user agnostic than a measure based on the in-game score.
Hb: This measure should constitute a proxy for the level of user control.

The experiment we conducted with users without motor disabilities has provided a positive answer to **RQ1.0** and **RQ1.1**. We used the game framerate as a parameter for balancing the lower user performance in their interaction with our optically tracked gamepad. The analysis of hand position throughput the experiment showed that users did adapt their behaviours with changes in the gamepad positions. However, we only observed a marginal increase in users motions, whereas the changes in the gamepad positions was four-fold.

Finally, **RQ1.2** was also answered positively. The purposefully designed measure of performance exhibited less variation across users than the game score. It also correlated with the game score and we found that when the measure of the game score was identical between the reference system and our system, this measure of performance was also identical. In other word, it could be used in place of the score for estimating whether a user behaves differently than the reference.

The preliminary experimentations we conducted in an hospital environment have validated that the motions elicited were satisfactory from the occupational therapists point of view. We also established that users with SCI could engage in game session with the system, provided the interaction area was customisable. More research is now needed to verify assumptions made about the setting of difficulty to maintain motivation.

Chapter 5

Eliciting Motions Through Positive Reinforcement

Summary. This chapter proposes to adapt the technique developed by Williamson et al. in the paper *Rewarding the Original* [107], which induces users into exploring the space of motions they can perform, to the particular case of upper limb interactions on tabletops. An updated version of the original algorithm is proposed and a random search over the parameter space is ran to find a combination suitable for the sensors and motions specific to upper limb interactions. A user study is conducted where the volume of user motions in unconstrained upper limb interactions is measured. This is used to draw comparisons with the motion spaces recorded in the two previous chapters. In particular, we find gesture typing motions to be more contrived than game induced motions. This exploration process of motions spaces, which relies on positive reinforcement through audio feedback, is then extended. From rewarding motion originality to explore variability, we seek to explore reliability in user motion space through the reward of motion repetitiveness. We show how a discriminative model can be trained to detects cycles in a continuous sensor stream. An experiment is designed making use of this model to elicit repetitive motions from participants.

5.1 Introduction

The previous chapter has presented the design of an upper limb gestural interaction for the purpose of the rehabilitation of the reach capability of users with SCI. During in-situ experimentations with users with severe motor impairments, see section 4.8, the need for a proper placement of the interactive surface and its orientation were identified as mandatory to match the region in which users could afford the motor control required by the interaction. The issue of matching user capabilities and interaction requirements abstracts the more fundamental problem of understanding user capabilities in context, which at minima encompasses the user motions and their sensing. It is the association of both which produces the material from which the interaction is constructed. *What type of motions, which are sensed by the system, can users produce?* A general approach to answering this question has already been proposed by Williamson et al. [107], where a process for systematically exploring the space created by sensed user motions was detailed. Recognising that it is "challenging to work out how to most effectively design new input systems, or how to best exploit the capabilities of existing devices", they propose "to consider how input mechanisms can be quantified in a broader sense, and how factors which influence the use of these mechanisms can be analysed within a coherent framework."

This chapter proposes first to summarise the formalism introduced by Williamson et al. before repeating their experiment in the context of upper limb gestural interactions created through optical sensing. Then, an extention to the technique is proposed targeting a property of motions that was not investigated in the original paper.

5.2 Joint User-sensor Space

In the very general case of a user engaged in an interaction with a computing device, the user's motions are sensed by the device's sensors, which outputs are used by the device to infer the user's intention, see 2.4. A feedback can be at that time conveyed back to the user, potentially closing the interaction loop. The space of user motions, as defined by the measurements made by sensors, is called in the following the *joint user-sensor space*. A joint user-sensor space depends on both individual users and device's sensors and is the material from which interactions are constructed. Two examples of very different interactions can help illustrate the interplay between users and sensors. During the interaction with an elevator button, a user has almost full freedom of movement, however the button's sensor only allows for a limited amount of information to be captured, mostly through pressure and on a rather small surface. On the other hand, a user engaged in the act of driving will find himself situationally impaired and only afforded with very limited movements. Yet, many complex buttons allow for a wide range of intentions to be transmitted. The amount of information that can be conveyed during the interaction depends thus on both user and sensor properties. Note that under this description, the purpose of the interaction is not significant, only the process of communicating intention is considered from an information point of view.

The analogy with a communication channel was used in the original paper. For a one-dimensional signal, the Shannon-Hartley equation computes the maximum amount of data that can be transferred in an analogue channel subject to additive white Gaussian noise by relating the bandwidth and signal to noise ratio to the throughput:

$$C = B \log_2 \left(1 + \frac{S}{N}\right)$$

giving a transfer rate of C bits per second from a bandwidth B and a signal-to-noise ratio S/N. The information throughput, or transfer of information, is proportional to the bandwidth of the channel and multiplied by the logarithm of the signal to noise ratio. For a signal to noise ratio tending towards zero, the information throughput also tends towards zero. Though simple, this model has been used successfully by Berdhal et al. [108] to compute the estimated channel capacity of an interaction between participants and different types of sensors.

Building on this analogy, two concepts related to properties of joint user-sensor spaces are introduced: *variability* and *repeatability*. Variability is analogous to the channel capacity, and in the joint user-sensor space it is the volume of possible sensed user motions. The level of variability depends both on the user and the sensor. The variability in terms of motions is subject to the user's capability to produce a large range of diverse motions, while the diversity in terms of readings is determined by the sensor's ability to produce a rich measurement from the observed user. Repeatability, on the other hand, relates to the signal to noise ratio. It is defined on both motions and readings and quantify the uncertainty in the signal. For a motion to be useful, it has to be controlled and produced on purpose.

Up to this point, motion spaces have been defined as joint user-sensor space, but no concrete definition has been provided for motions themselves. Here again, a clear distinction is made between information and semantics, equivalent to the difference between motions and gestures. Only the sensor output is used as the basis for defining motions as the sampled and discrete version of sensed user motions. Time derivatives of these readings, in lieu of time stamps, are included as to provide a notion of evolution over time. This is the description that was used by Williamson et al. [107] and is equivalent to definitions popular in the field of robotics [109].

Mathematical Formalism

On a more practical level, a mathematical formalism can be used to define joint user-sensor spaces and associated sensed motions. A joint user-sensor space is composed by the set of samples from the sensor readings of measured user motions and associated derivatives. Let each sample be a D-dimensional vector $\mathbf{x}$:

$$\mathbf{x} = (x_0^0, x_0^1, \ldots, x_0^m, x_1^0, \ldots, x_n^m), (m, n) \in \mathbb{Z}^+$$

from a sensor producing n-dimensional output with m associated derivatives, and $D = m \times n$.

The number of dimensions depends on the type of sensing used in the system. Images usually have thousands of dimensions: one for each pixels and additional ones for each channels in each pixels. Inertial measurement units (IMU) on the other hand output data with typically tens of dimensions. For instance, the original paper was designed to capture the motions produced by a joint user-sensor system composed of a human upper limb strapped with one IMU, which produced readings from its accelerometer and gyroscope sensors measuring linear and rotational accelerations, each being 3-dimensional. Such data was then submitted to analysis (low-pass filtering, time resampling and interpolation) in order to compute the first three derivatives of the sensor readings. The resulting data stream is a vector with 36 dimensions producing a new observation 32 times per second.

By considering now $\mathbf{x}$ as a D-dimensional random vector, with each dimensions being a scalar random variable $\mathbf{X_i}$, we can define the joint probability distribution which models the relationships between dimensions. The covariance matrix of this distribution can be measured with:

$$cov[\mathbf{x}] = \mathbb{E}[(\mathbf{x} - \mathbb{E}[\mathbf{x}])(\mathbf{x} - \mathbb{E}[\mathbf{x}])^T]$$

and is defined to be the following symmetric, positive definite matrix:

$$cov[\mathbf{x}] = \begin{pmatrix} var[\mathbf{X_1}] & cov[\mathbf{X_1},\mathbf{X_2}] & \dots & cov[\mathbf{X_1},\mathbf{X_d}] \\ cov[\mathbf{X_2},\mathbf{X_1}] & var[\mathbf{X_2}] & \dots & cov[\mathbf{X_2},\mathbf{X_d}] \\ \vdots & \vdots & \ddots & \vdots \\ cov[\mathbf{X_d},\mathbf{X_1}] & cov[\mathbf{X_d},\mathbf{X_2}] & \dots & var[\mathbf{X_d}] \end{pmatrix}$$

with the variance of $\mathbf{X_i}$, noted $var[\mathbf{X_i}]$, defined by:

$$var[\mathbf{X_i}] = \mathbb{E}[(\mathbf{X_i} - \mathbb{E}[\mathbf{X_i}])^2]$$

The covariance matrix gives access to interesting properties of the joint distribution. The covariance between pairs of dimensions gives an indication of structure in the motion space. The eigen vectors of the matrix follow the principal directions of the joint user-sensor space, while the eigen values provide a measure of scale along these directions. As a result, an estimate for the total volume $V_\mathbf{x}$ of the space that is being described by such distribution can be accessed through the determinant of the covariance matrix:

$$V_\mathbf{x} \propto det(cov[\mathbf{x}])$$

and can be compared to the volume produced by the distribution of another random vector.

A measure of distance can also be defined on this vector space. To take into account the difference in variability between dimensions, the Mahalanobis distance [57] can be used. It is defined between two vector $\mathbf{x}$ and $\mathbf{y}$ sampled from a distribution with covariance cov as:

$$D_M(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})^T cov^{-1} (\mathbf{x} - \mathbf{y})}$$

The Mahalanobis distance D_M produces a measure that is unit-less and scale-invariant. The measure of distance between two vectors gives access to their dissimilarity or to the originality of one vector as compared to a collection of other vectors. This was used in the original paper as the measure of originality between samples taken from the joint user-sensor space.

In this section, the joint user-sensor space as been defined as a D-dimensional random vector space, with its joint distribution and covariance matrix, and a metrics that can be used to assess originality between two vectors has been introduced as the Mahalanobis distance. The following section will detail the process through which joint user-sensor spaces can be explored and characterised.

5.3 Automated Elicitation as a Search Problem

Before continuing with the rest of this chapter, it is opportune to make here a parallel between the approach we propose and another type of well-established interactions: elicitations studies.

Elicitation are techniques which seek to gather knowledge directly from users [110]. Elicitation studies in HCI are usually aimed at associating motions and meaning by discovering gestures that are guessable [36]. This relies on two fundamental parts: a search process performed by users and a measure that determines the outcome of a selection. The procedure usually involves a prompt by the experimenter who chooses a referent for which users are supposed to perform a gesture. The internal process of search followed by participants is captured as they think aloud. Then a measure of agreement on the proposed gestures is computed by analysing the result of associations participants make between proposed gestures and prompted referents.

In the original paper, joint user-sensor spaces have been used in a similar fashion, with some notable differences. In lieu of gestures, motions were produced and instead of the guessable nature of gestures, the variability of motions was sought after. The process of search relied on placing users in a reinforcement learning scenario where rewards would be provided every time a new motion satisfied a condition. To emphasise variability, the novelty of an observed motion was computed as the Mahalanobis distance of the sample representing it to the joint user-sensor space distribution that had been captured until then.

By making a connection between both type of studies, we see that our approach can be described as a data-driven (or automated) elicitation technique, where quantitative and arbitrary properties of motions can be sought after, by placing users in a reinforcement learning scenario. This search process also can be described by drawing from the field of Artificial Intelligence [111]. This, for instance, introduces the interplay between exploration and exploitation. As an agent is performing a search it must balance two opposite objectives which are to explore the totality of the searchable space while exploring it thoroughly. The rest of this chapter will build on this idea and present two application for this, by repeating the original experiment from Williamson et al. which looks for variability and by proposing a variant which targets repeatability.

5.4 Variability in Joint User-sensor Space

A process, named "Rewarding the Originial" (RTO), was designed to explore the variability in joint user-sensor space. A joint user-sensor space is first chosen. For example, a body part such as the user arm is selected and associated with a sensor such as an optical tracking system.

The goal of the process is to elicit variability in joint user-sensor spaces. In other words, the goal is to map the totality of the motions that can be both produced by users and sensed by the system's sensors. Users are placed in a reinforcement learning process in which they search for originality in their produced motions through both exploration and exploitation. The search is informed by a reward proportional to the originality of the current observed sample as computed by the Mahalanobis distance with previously observed samples. For clarity, we will call *observation* a sample from the joint user-sensor space that has been deemed original and has thus triggered a positive feedback, and *catalogue* the set of *observations*. The purpose of the RTO process is to incrementally record the catalogue. A sample is declared an observation when the average value of the Mahalanobis distance of the sample with the K closest observations from the current catalogue is greater than a threshold T. After each new observation is stored, the covariance of the catalogue should be updated.

The independent variables or parameters of this process are the threshold T, the number of neighbours K and N the number of new observations between a re-computation of the covariance of the catalogue. Note that the covariance of the catalogue should be re-computed each time a new observation is added, but is in practise computed every N new times for processing cost reasons. The dependent measures registered by this process are the catalogue of observations with their time of discovery or recording, from which the interaction volume over time can be computed.

Algorithm and Parameters

For the purpose of running RTO with upper limb gestural interactions on planar surfaces, the RTO algorithm was reimplemented with a small modification. One issue with the original algorithm, discovered during early experimentations, was that samples both very similar and adjacent in time would be qualified as observations. This was due to the averaging of the Mahalanobis distance with K neighbours and the intermittent computation of the covariance of the catalogue. To ensure a more uniform sampling, another condition on the novelty detection was added in the form of a minimum threshold to the distance to the closest observation in the catalogue. In other word, the threshold T was replaced with a combination of two thresholds D_{mean} and D_{min}. The algorithm's pseudo-code is detailed below:

Algorithm 1 Modified RTO

Require: D_{mean}, D_{min}, K, N
  $catalogue$ = []
  while new $\mathbf{x}$ **do**
    $distances$ = []
    for all $\mathbf{y}$ in $catalogue$ **do**
      $distances.insert(D_M(\mathbf{x}, \mathbf{y}))$
    end for
    $distances_{mean} \leftarrow mean(sort(distances[: K]))$
    $distances_{min} \leftarrow min(distances)$
    if $(distances_{mean} > D_{mean})$ **and** $(distances_{min} > D_{min})$ **then**
      $catalogue.insert(\mathbf{x})$
    end if
    $counter \leftarrow counter + 1$
    if $mod(counter) = N$ **then**
      recompute $cov(catalogue)$
    end if
  end while

The value for these parameters is arbitrary and were empirically set in the original paper. Good values for these parameters allow for a reward that is not too frequent to avoid collecting each and every sample, and not too seldom to avoid discouraging participants in their exploration process. A small experiment was thus conducted to find appropriate values for the specific case of upper limb gestural interaction on planar surfaces.

Specific case of Upper Limb Gestural Interactions on Planar Surfaces

The sensor used for this experiment was an Intel Realsense SR300 depth camera. The same simple tracking algorithm, developed for the previous chapter, see 4.8, was used. This tracker can follow one reflective marker at $30Hz$. The camera was overlooking an office desk, placed in such a way that regions reachable by a user were fully covered.

This original vector stream is uniformly re-sampled at $30Hz$ and submitted to additional filtering. A Savitsky-Golay interpolator on a $500ms$ window with a 3rd order polynomial computes the 1st and 2nd derivatives of motion. We have thus 6 spatial dimensions: position, velocity and acceleration on a 2 dimensional plane. In the following sections, the dimension are marked with the letter X or Y reflecting horizontal and vertical dimension on the tabletop, respectively. These letters are appended a number between 0 and 2 reflecting the derivative number, see Table 5.1.

	Horizontal	Vertical
Position	X0	Y0
Velocity	X1	Y1
Acceleration	X2	Y2

Table 5.1: Naming and signification of the d-dimension of the joint user-sensor space.

A computer ran the RTO algorithm in real time. The value of the originality measure was displayed as a curve on a screen facing the interaction area. The task for participants was to maximise the value for the originality measure over a period of 10 minutes. One participant, the author, took part in the experiment. A reflective marker placed between the index and middle finger was used to indicate the hand position and was attached with a stretchable strap. We collected 18184 samples from this joint user-sensor space.

The dependent variable we were interested in is the proportion of samples that would be considered as new observations and added to the catalogue. We ran a random search on the recorded data to estimate the distribution of this proportion as a function of D_{mean} and D_{min}. Note that the value for K, the number of neighbours, was set to 5, and the period at which the covariance was recomputed was also set to 5. The result are plotted on Figure 5.1. We observed a decrease in the proportion of samples added to the catalogue with an increase of D_{mean} and D_{min}, with 90% of the variation explained for values of D_{mean} and D_{min} comprised between 0 and 2.

We arbitrarily chose a combination for D_{mean} and D_{min} that produced 15% of rewards on the training dataset, see Table 5.2, similar to the original experiment.

Parameter	Value
D_{mean}	1.33
D_{min}	0.60

Table 5.2: RTO parameter values for D_{mean} and D_{min}.

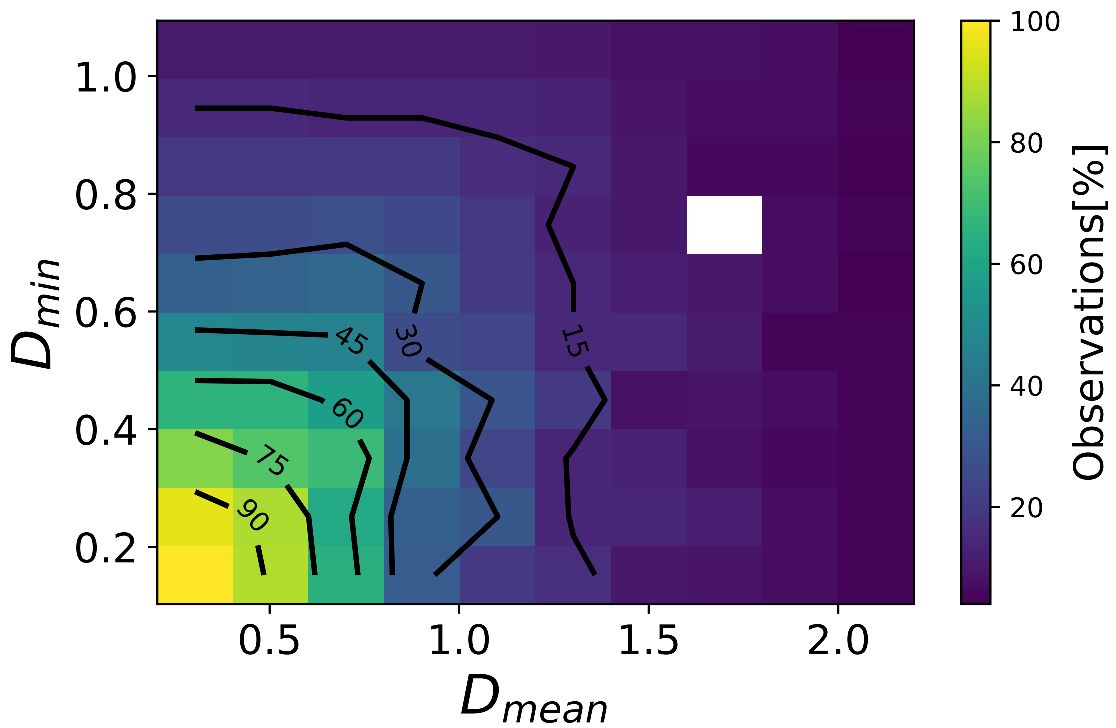

Figure 5.1: Percentage of samples stored in the catalogue as a function of D_{mean} and D_{min}. Note that the white square in the top right is due to an absence of sampling data.

5.5 RTO Experiment

We designed an experiment to capture the joint user-sensor space for the specific case of upper limb gestural interactions on planar surfaces. This constitutes a generalisation of the interactions that have been implemented in the two previous chapters, since motions are unconstrained by a task such as text-input or pointing. The only constraint is the process that engages participants in a positive reinforcement loop.

5.5.1 Apparatus

The same optical tracking system as in section 5.4 was used in this experiment. A reflective marker placed between the index and middle finger was used to indicate the hand position and was attached with a stretchable strap. The interaction area provided by this setup was a square of dimensions $78cm$ by $78cm$. It was designed to exceed the participant's reach in the forward direction when participants had their trunk kept immobile and in contact with the chair they were seated on.

A computer ran the RTO algorithm and provided an audio feedback each time an observed sample was added to the catalogue of original motions. The audio feedback was designed as a constant sine wave with frequency at $90Hz$ which was smoothly pitched upwards by another $90Hz$ for every new detection of an observation. This positive feedback had a linear decay lasting $200ms$. The resulting feedback indicating originality was a short upward pitch from a constant low background noise.

5.5.2 Task

The task for the participant was to produce motions that would trigger the audio feedback indicative of the originality of the current sensed sample.

5.5.3 Procedure

Participants were equipped with the reflective marker on their dominant hand, in between the index and middle finger and attached with a stretchable band. Participants were placed on a chair by a tabletop and instructed to have their back in contact with the back of the chair at all times. The chair was placed to the side of the tabletop while staying within the tabletop's legs in order to have the dominant arm of the participant placed in the centre of the tabletop. The chair was also positioned so as the elbow of the participants would comfortably rest on the closest edge of the tabletop. The centre of the interactive region was placed one full forearm extension towards the camera from the edge of the tabletop.

Once in place, participants were informed that their motions would provide audio feedback based on the position, velocity and acceleration of their hand and that their task was to maximise the increase in pitch in the audio feedback. They were subsequently equipped with a headset that was conveying the audio feedback.

5.5.4 Design

All participants were subject to the same condition in this experiment. The task was designed to run for five minutes, after which the experiment was concluded.

5.5.5 Participants

We recruited six participants, including two female participants, with a mean age of 35.5 and standard deviation of 11.2 (min=23, max=55). All participants were right handed. The study was approved by the University of Glasgow ethics committee. No participants presented any mental or physical disabilities.

5.5.6 Results

The data collected through the experiment were the vectors of the user-sensor space, the original observation of the catalogue with their time of discovery and the value of the reward for each samples. After analysis of the logs, we realised that the data for one participant was compromised. The following analysis is carried with the remaining five participants only.

Distributions of Position, Speed and Acceleration

The position data is plotted on Figure 5.2 with summary statistics on Table 5.3.

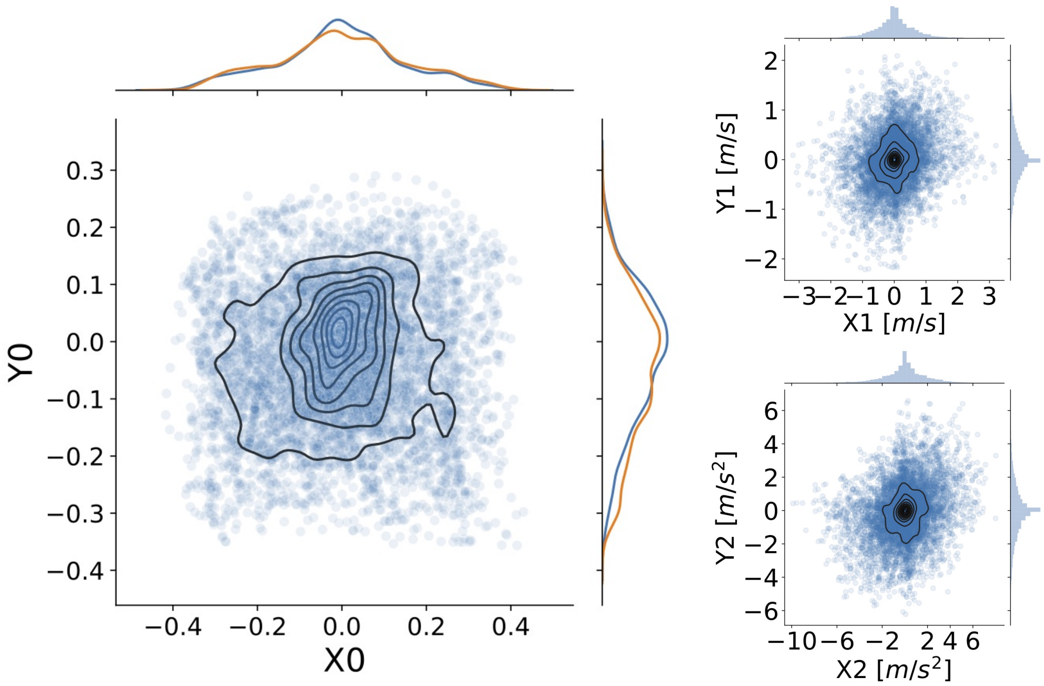

Figure 5.2: Joint probability distribution of the hand position, speed and acceleration during RTO.

	X0	Y0	$[m]$	X1	Y1	$[m/s]$	X2	Y2	$[m/s^2]$
mean	-0.00	-0.03		0.01	-0.01		-0.02	-0.06	
std	0.15	0.11		0.68	0.53		2.01	1.62	
min	-0.42	-0.36		-3.19	-2.22		-9.88	-6.22	
max	0.43	0.29		3.20	2.09		8.03	6.58	

Table 5.3: Summary statistics for the D dimensions of the joint user-sensor space.

A scatter plot with low transparency shows the totality of the catalogue in X0 and Y0 for all participants. A joint distribution was computed on top of the scatter plot. The joint distribution exhibits a higher density around the origin (0,0) indicating that participants did adopt their starting position (back in contact with the chair and elbow on the edge on the table) as their rest position. This is corroborated by the mean value in X0 and Y0 at 0.0. The joint distribution displays a higher density in the lower left corner, representative of the flexion of the arm for right-handed participants. The marginal distributions in X0 and Y0 were also computed. The blue and orange curves represent the marginal distribution for the full dataset and the catalogue only, respectively. The marginal distribution in X0 is almost symmetrical and extends to the full limits of the tabletop with a minimum and maximum at roughly $40cm$. The marginal distribution in Y0 is slightly skewed towards negative values reflecting the limit in reach of our participants. The maximum Y0 value is at $29cm$ as compared to a minimum value of negative $36cm$.

Similar jointplots were computed for the first and second derivative of motions (X1, Y1) and (X2, Y2). Both derivatives exhibit symmetrical distributions with a sharp peak at the origin. The maximum values for the horizontal and vertical velocities lie at $3m/s$ and $2m/s$, respectively. In a similar way to the position data with bigger horizontal range, horizontal motions are also favoured in terms of velocity. Acceleration data presents the same characteristics, with maximum positive horizontal acceleration at $10m/s^2$ and maximum negative horizontal acceleration, or deceleration, at $8m/s^2$. This slight imbalance might also find its origin in the asymmetry of the upper limb, with different properties for the muscles responsible of the flexion and extension of the arm. On the other hand, the vertical acceleration is symmetric and shows a maximum norm around $6m/s^2$.

Distributions of Maximum Speed and Acceleration

To display the relationships between the position and its derivatives velocity and acceleration, we computed the norms of velocity (speed) and acceleration (scalar acceleration), as the norm of the vectors composed by the horizontal and vertical derivative components. We stored the maximum value of speed and scalar acceleration for different value of the position organised in 30 bins. The results of this operation are plotted on Figure 5.3 and Figure 5.4 for four participants, allowing for a qualitative comparison between the participants' derivative profiles. For the speed profile we observe a concentration of high speeds in the middle of the explored region with a constant decreasing trend as the position reaches the limit of participants' reach. It is worthy noting that the distribution of maximum speeds differs across participants, most notably for the participant plotted in the lower right corner with much lower maximum speeds than the average. For the scalar acceleration profile we obtained a slightly different picture, with more pronounced non-uniformity in the distributions. We noted the presence of clusters of high accelerations on the outer bounds of the position with lower values in the middle of the tabletop.

These plots, which characterise the capabilities of individual subjects, are one piece of information for answering the question that was asked in the introduction. *What type of motions,*

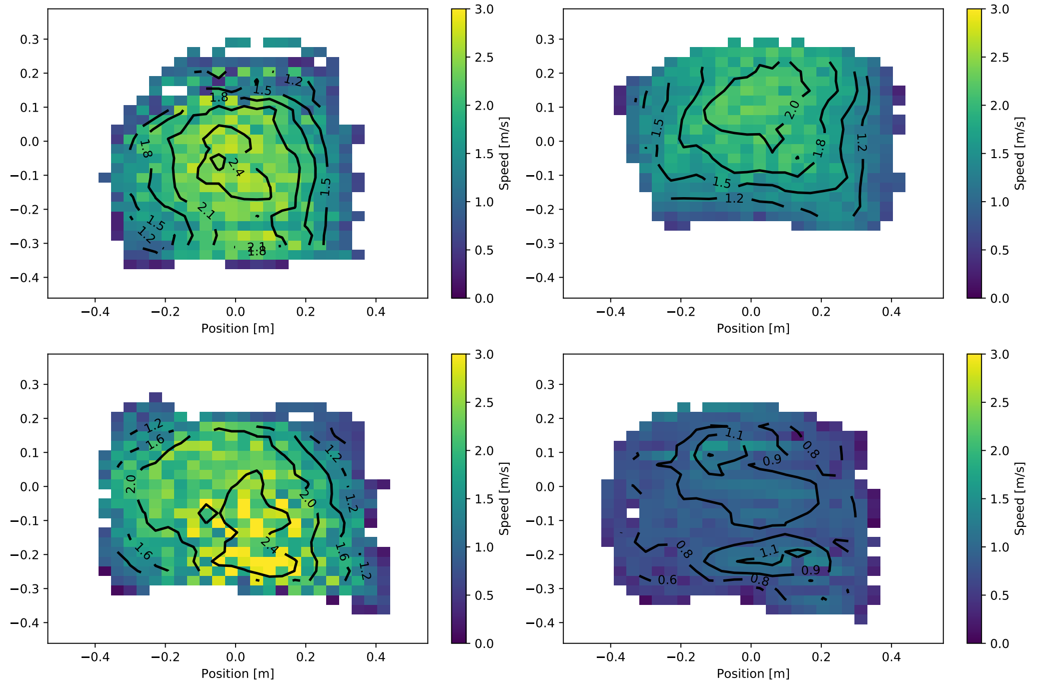

Figure 5.3: Maximum value of the speed as a function of the position, for four participants.

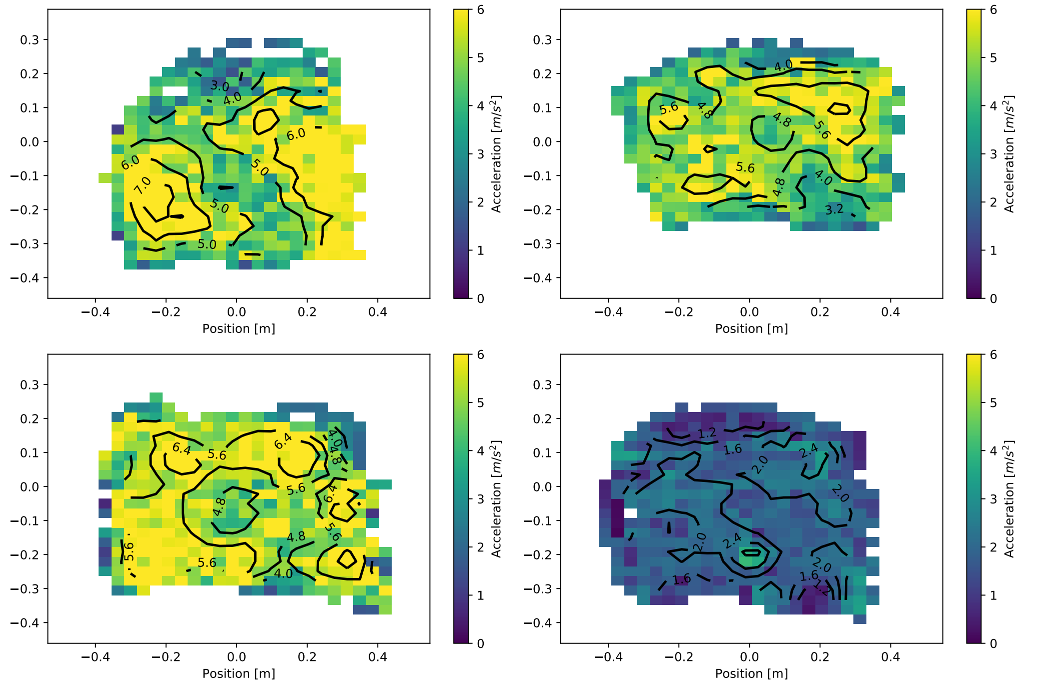

Figure 5.4: Maximum value of the scalar acceleration as a function of the position, for four participants.

which are sensed by the system, can users produce? They inform about the reaching capability of users and also about the dynamics of observed user motions, which could be related not only to their static abilities but also to a patient's ability to exert force through the measure of velocity and acceleration. As a result, there is some potential for using this tool in the context of rehabilitation and for OTs to access a quantitative measure of their patients' abilities, or as a complement to a more qualitative measure of reachable space, such as the one proposed by Toney et al. [7].

Note also that due to the nature of the sensors, which provide an absolute measure of the hand position and access to its derivative, the maps of Figure 5.3 and Figure 5.4 are much more interpretable than if we were using accelerometer data. Indeed, these produce readings that are second derivatives of the motions that generated them, making them harder to link back to what a human could see without integration. As a result, the data we collected in the experiment was easier to relate to the reaching capabilities of a user, as well as its maximum capability in terms of speed and acceleration.

Volume of Captured Space

Beyond the analysis of individual properties of the D dimensions of the vectors from the catalogue, the analysis of the volume of the *catalogue*, $V_{catalogue}$, and its evolution over time provides additional information about how participants reacted to the RTO process.

The timestamps of *observations*, each having produced a positive reward during the experiment and added to the *catalogue*, were analysed. The cumulative number of observations as a function of experiment time is plotted on Figure 5.5 (left), with associated values in Table 5.4. The different colours represent our five participants and are kept across both figures. The average value of saved vectors in the catalogue is 1525, with a minimum at 804, a maximum at 2087 and a standard variation at 494. We observe that the rate at which participants did receive rewards varies across participants and across time. For example, the red and blue curves present a continuous smooth increase throughout, while the other curves present clearer discontinuities which appear for the green one around minute 1, for the orange one around minute 2 and for the purple one around minute 3. These discontinuities can be explained by the sudden discovery of another existing type of motions that participants did not propose until then, leading into the exploitation of that new type of motions. For instance, such motions could be high accelerations and deceleration or farther reaching motions. These discontinuities reflect the alternative nature of exploration and exploitation during the systematic search of the joint user-sensor space conducted by the participants.

The number of observations is only one side of the picture. The volume of the distribution of observations was also computed as a function of experiment time. Note that we used the logarithm of the volume for comparison purposes. It is plotted on Figure 5.5 (right), with summary statistics in Table 5.4. The average value for the volume explored is -9.53 with a minimum value at -14.27 and a maximum value at -5.88. Overall, all curves present an

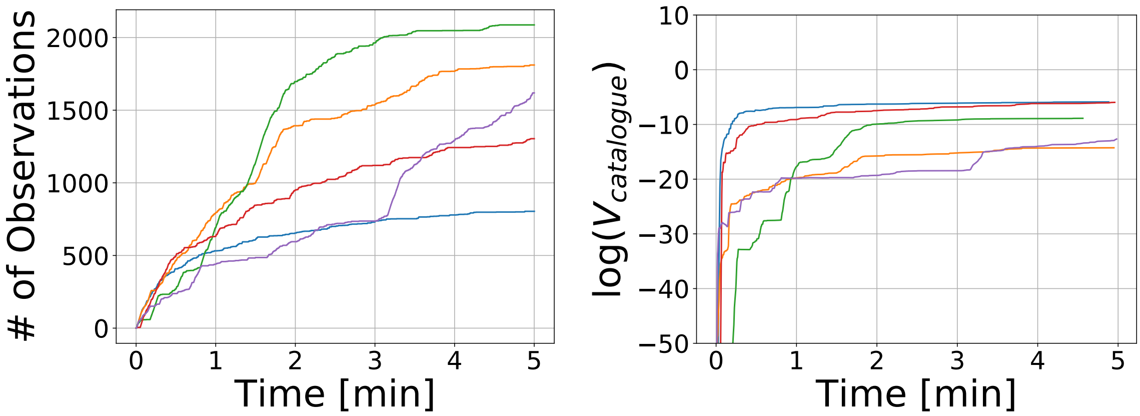

Figure 5.5: Number of vectors present in the catalogue (left) and volume of the catalogue (right) as a function of experiment time.

	# of Observations	$\log(V_{catalogue})$
mean	1525	-9.53
std	494	3.83
min	804	-14.27
max	2087	-5.88

Table 5.4: Summary statistics for the number of observations and $log(V_{catalogue})$ at the end of the task averaged over participants.

asymptotic behaviour with a limit that seem to have been closely approached for all participants, indicating that the rate of rewards and the length of the task have allowed participants to explore the majority of their joint user-sensor space. The curves present a similar evolution to their counterpart in terms of number of observations, when considered for the same user. However, it is important to note that there is no direct correlation between the number of observations and the volume of catalogue, as perfectly illustrated by the blue and green curves. In other words, while the participant represented by the green colour did earn the maximum number of rewards during the experiment, the associated volume is smaller than the one uncovered by the participant represented with the blue colour, who also happened to generate the least amount of rewards during the experiment. This difference is explained by the different exploration strategies employed by both participants. The blue participant did use a conservative exploration method with small steps, generating a lot of incremental rewards, while the green participant did employ a more aggressive exploration strategy involving a higher and earlier diversity of motions. The influence of the design of the audio feedback could come here into play. The audio feedback did convey a binary measure of originality, whereas it could have also been modulated by its absolute value.

5.5.7 Comparison with other Upper limb Interactions

One of the benefit of the measure of volume in the joint user-sensor space is that is allows for meaningful comparison between spaces. We have thus computed the same metrics for the dataset that were collected in the previous two chapters, namely the interaction involving gesture typing and the interaction dedicated to games in rehabilitation. The volume across different conditions is presented on Figure 5.6, see section 3.5.3 and section 4.5.3 for the significance of the testing conditions. We observe that the volume discovered in the RTO experiment is bigger on average than any other volumes. For the data from chapter 3, we observe an increase in interaction volume with an increase in the interaction size. For the data from chapter 4, we observe an increase of the interaction volume with an increase in T_RATE and SPREAD. The range of hand positions and level of velocity and accelerations involved are likely responsible for this effect.

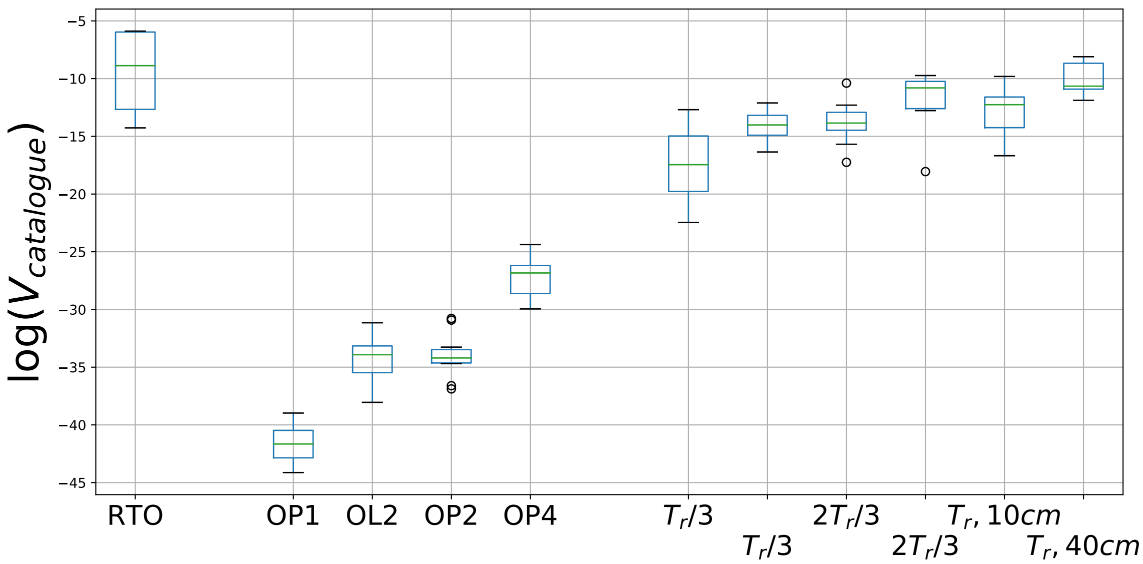

Figure 5.6: Volume of explored joint user-sensor spaces for the RTO experiment (first box), the gesture typing experiment from chapter 3 (group in the middle) and the game experiment from chapter 4 (group on the right).

Next, we were interested in the relation between interaction volume and information throughput that was afforded by the interactions from Chapters 3 and 4.

The information throughput from the gesture typing data is hard to compute due to complexity of decoding technique. However, the results from the experiment indicated that the input rate measured in word per minutes was equivalent across conditions. This signifies that the interaction with OP1 managed to make a better use of its space than the interaction with OP4: with a much smaller volume, the information throughput is maintained.

For the data from Chapter 4, however, it is possible to easily compute the information throughput reached during the interaction. The user was capable of selecting 4 different virtual buttons via a specifically designed gamepad, each encoding one action that was sent further to the game it was connected to. From Information Theory, we know that the number of bits encoded by selecting one action when N actions are possible and equiprobable is given by $log_2(N)$, which here equates to $2bits$ of information per selection. Thus, we computed the information throughput as the number of action selection multiplied by $2bits$ and divided by the time over which the selections occurred. The results are presented in Figure 5.7. The plot represents the information throughput IT as a function of the volume. A linear regression shows an effect ($p < 0.001$), indicative of a non-zero slope, with parameters ($slope = 0.06$, $intercept = 2.13$, $stderr = 0.014$). However, R-squared is low with a value of 0.22, indicating that most of the variance in the data is not explained by the linear relationship between the information throughput and the volume used during the interaction.

What we can take away from these results is that the interaction volume does not correlate well with the information throughput reached during the interaction. This discrepancy will be discussed in more detailed in the following section where we discuss the importance of not only variability but also repeatability in joint user-sensor space for a complete description of the information transfer.

5.5.8 Conclusion

The RTO algorithm has been adapted to the special case of upper limb gestural interactions on planar surfaces. The value of its parameters was informed through experimentation with training data. A user study was designed to capture the properties of joint user-sensor spaces of users in the context of our category of interactions. We have shown that there exists some structure in the dataset, with non-uniform distributions of positions, speeds and accelerations of users' end effector. The data has also uncovered some variance between participants, most notably with one user exhibiting lower derivatives of hand positions.

We have compared different input modalities and drawn conclusions about the volume different interaction were putting into action. The interaction volume and the information throughput did not show a strong linear relationship.

5.6 Repeatability in Joint User-sensor Space

We introduced the concept of repeatability in the introduction via the analogy with the notion of signal to noise ratio from information theory, whereas variability was associated with the bandwidth of the channel. The lack of strong relationship between the interaction volume

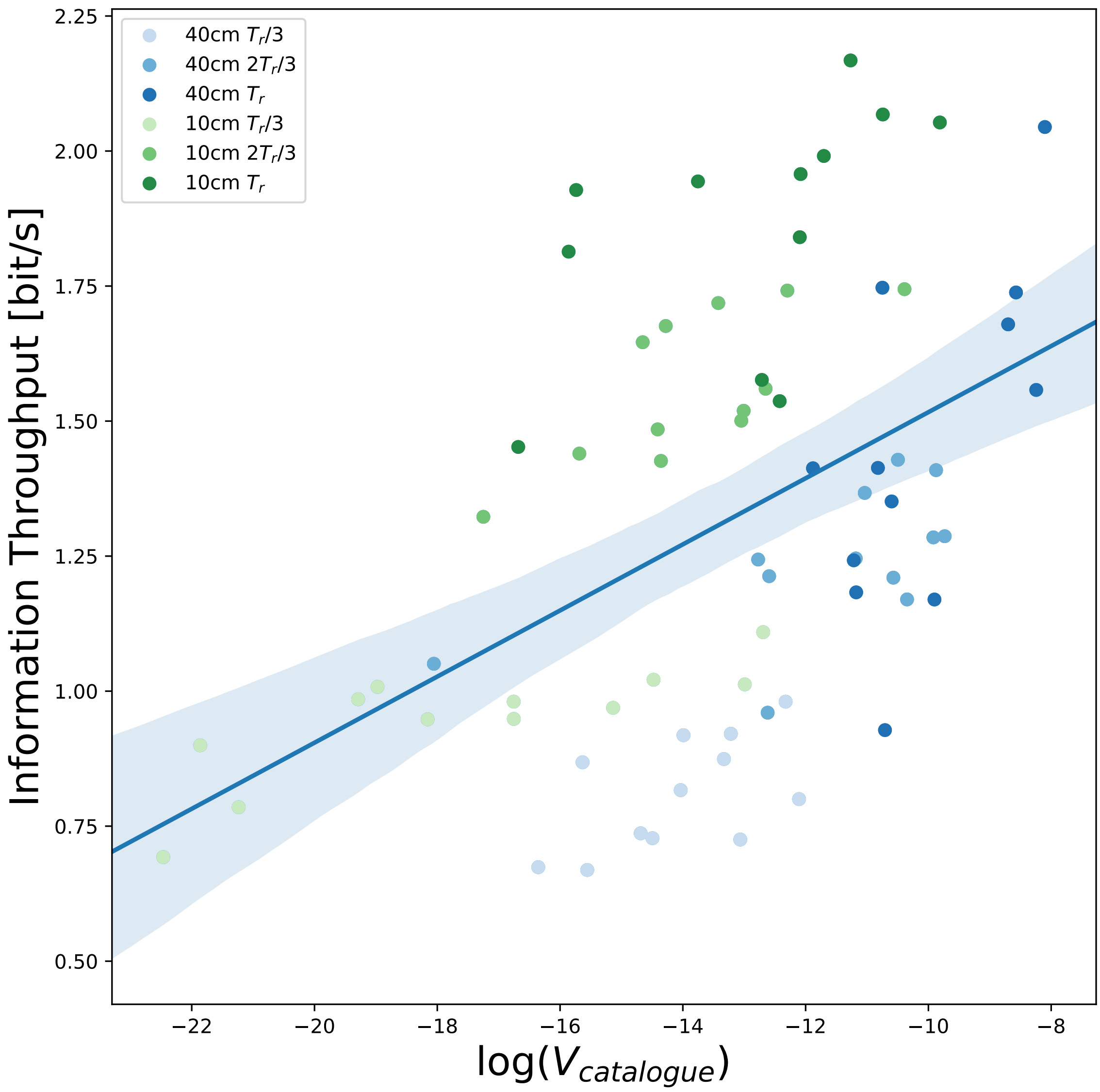

Figure 5.7: Information throughput as a function of volume for the six conditions of the user study in Chapter 4.

and the information throughput resides in the fact that repeatability was not taken into account. The remainder of this chapter is thus investigating the measure of repeatability in joint user-sensor space. There exists a relationship between information throughput and repeatability. Information throughput has been measured in Fitts' experiments for pointing tasks, or by Berdahl et al. [108] in their experiment on tracking tasks. For more complex gestures however, the notion of target is hard to express. Oulasvirta et al. [59] found a work-around by asking human subjects to repeat predefined complex trajectories. A professional dancer was recruited and was asked to perform a well-known routine two times in succession. This was used to model differences between repetitions, or repeatability in their motions. Here, we propose to combine the notion of repetitions of motions to the concept of positive reinforcement that was central to RTO.

The goal of creating a reward for reinforcement poses the problem of qualifying repetitiveness in observed motions, usually followed with the question of the segmentation of motions. The recognition, segmentation and classification of temporal data is an active field of research and several techniques have been proven successful for offline analysis. Lu et al. [112] have proposed to use autoregressive models for fitting the time series of human joint angles and positions, inferring repetitiveness based on the model's parameters stability over time. Morris et al. [113] used hand-crafted features and machine learning techniques for manually annotated data captured by accelerometers. Kruger et al. [114] computed a similarity matrix and detected frame of similarity and repetitiveness by analysing its structure. With a technique available in real-time albeit on image data, Levy et al. [115], have used synthetic data generation and neural networks to automatically detect and count repetitions in videos.

In this work, we propose to combine ideas from these different techniques. The need for live recognition points towards the method used by Levy et al [115] with synthetic data and neural networks. The idea of structure in the similarity matrix, which is a 2-dimensional computation of distance, will be adapted to the Mahalanobis distance which has been used until now in joint user-sensor spaces.

5.6.1 Generating Synthetic Motions

The notion of repeatability in motions is hard to define. Supervised learning has however proven successful for classification tasks, provided a labelled dataset is available. We create such dataset via two simple models producing motions we expect to observe from participants. The first one models random motions which are not periodic, whereas the second model is aiming at producing cyclical noisy motions.

Random Motions

Random motions are modelled as a low-pass filtered random walk without accumulation. The steps are generated from a 2-dimensional Gaussian distribution, with classical parameters location and scale, that is repetitively sampled. Each new sample represents a new position on a 2-dimensional plane (in X and Y) that is subsequently low-pass filtered. The values for location and scale are taken from two uniform distribution with support over (-0.5, 0.5) and (0.1, 1), respectively. Since the cut-off frequency of the low-pass filter is left unchanged, the scale parameters effectively model different movement speeds. Finally, for each sample of the pair location and scale, the resulting walk is a random sequence that can be observed for an arbitrary numbers of steps. Figure 5.8 shows a slice of the data that was

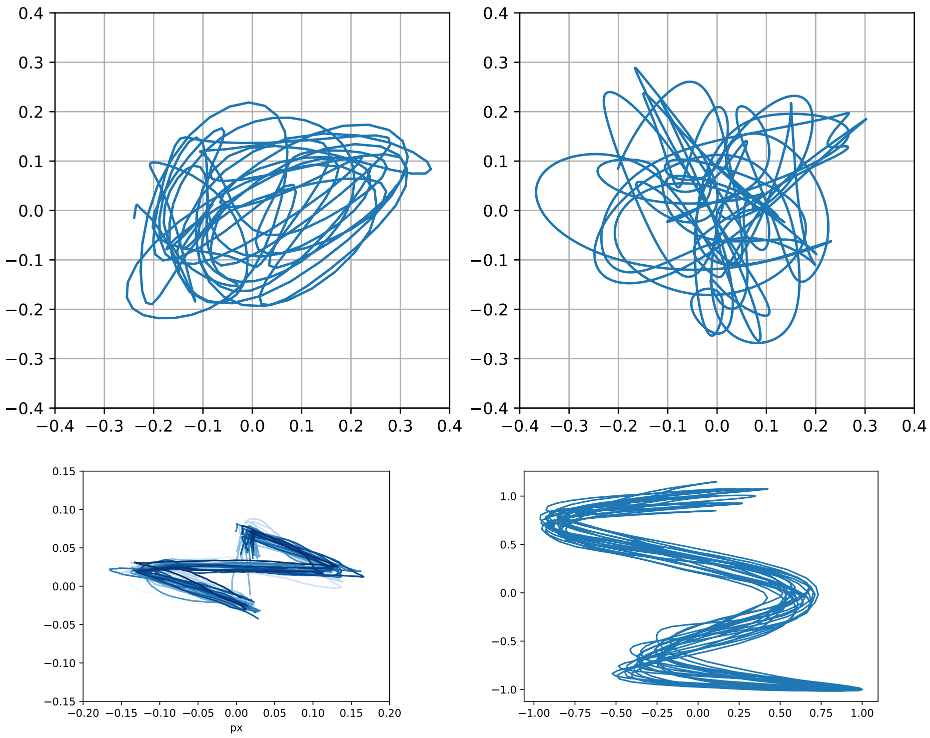

Figure 5.8: Comparison between real and synthetic motions. On the top row, selected time series of motions observed in the RTO experiment and generated by random walk, on the left and right, respectively. On the bottom row, motions taken from the gesture typing experiment and generated by a random oscillator, on the left and right, respectively.

recorded in the RTO experiment, see 5.5, side-by-side to a random sample generated by the above process for comparison.

Periodic motions

The generation of random repetitive movements is based on the modelling of a motion as a noisy oscillator. Fourier analysis states that any movement can be decomposed as a superposition of periodic functions. Repetitive motions can thus also be modelled as a sum of oscillators. The generation of a cycle is based on using the Discrete Fourier Transform (DFT) in real space. Because the DFT is defined for a periodic function, the inverse of the DFT (IDFT) generates a periodic function. To create random periodic motions, we randomly populate the spectrum of an oscillator before taking the IDFT, which generates a full oscillation. The noise component is then added to this oscillation. To model the human noise in executing the repetition, we chose to add some Perlin noise [116], as a function of the position and the time. The maximum noise amplitude is chosen to be 30% of the maximum amplitude in cyclical repetitions. The results produced by this process are shown on Figure 5.9 for 16 random samples. For comparison purposes, Figure 5.8 shows a motion taken from Chapter 3 during the gesture typing task. Even if this gesture typing motions is not periodic, some randomness can be observed along the repeated trajectories.

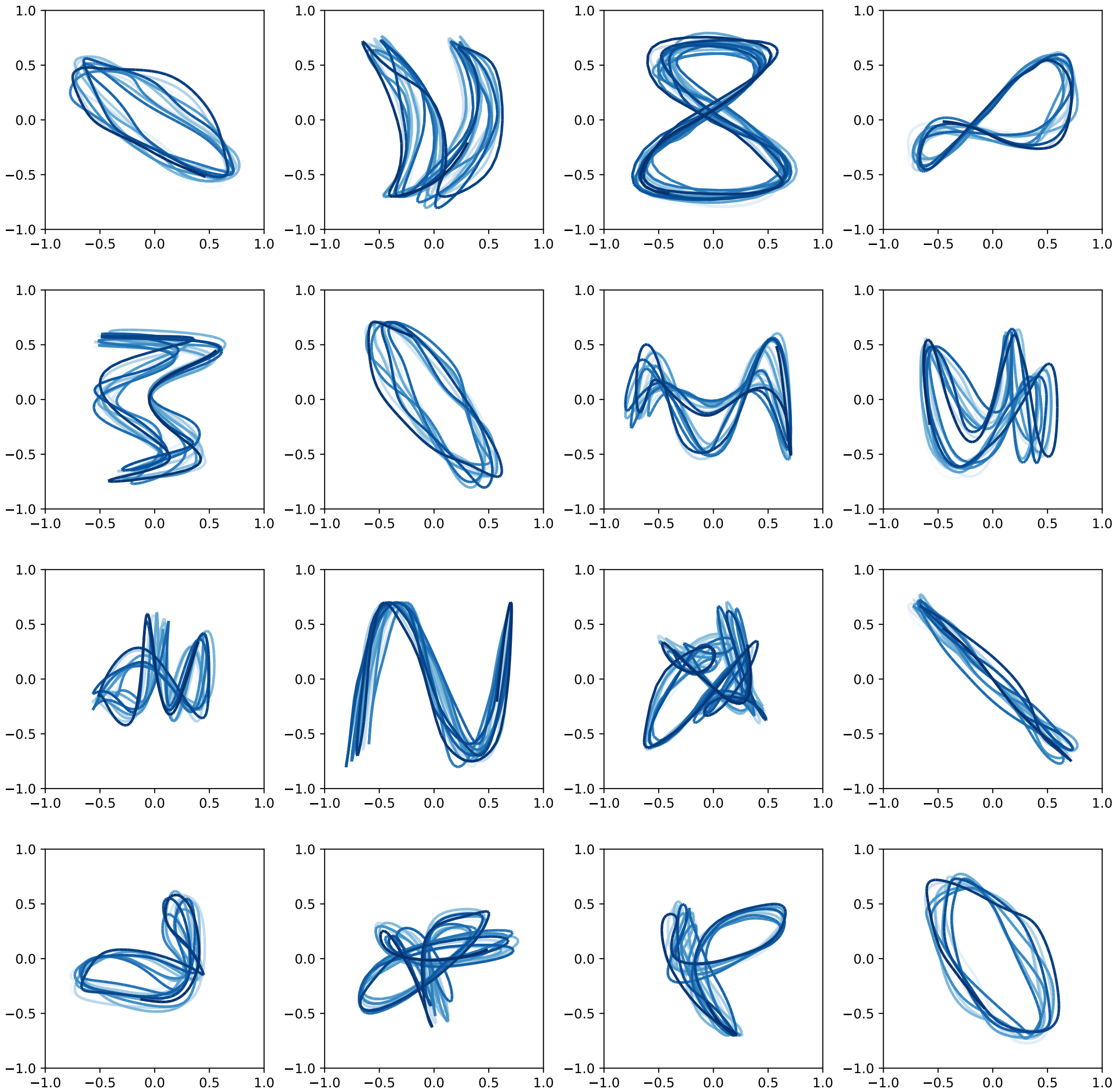

Figure 5.9: 16 samples from the random oscillation generation, with 30% of Perlin noise.

A labelled dataset of arbitrary size can be generated by continuously sampling these two models.

5.6.2 Detecting Repetitions

One of the conditions for a successful outcome of a classification task is to provide salient information to the model which is making the inference. This information, in the form of features, can be engineered or it can be extracted through a process that is also learned. We looked at the evolution in time of the Mahalanobis distance for the two categories of motions from our synthetic dataset. For each motions, the derivatives are computed so as to obtain the 6-dimensional data type we have used in our joint user-sensor space. The consecutive samples are stored in a FIFO buffer of length 120 samples and the Mahalanobis distance is computed on this buffer at each new sample. The result of this computation is also stored in a bi-dimensional FIFO buffer of dimension 120 by 50. The content of this buffer and the motion that was used for its computation are plotted side-by-side on Figure 5.10 and Figure 5.11 for cyclical and random motions.

What appears in this buffer is a curve that unfold in two dimensions. On the x-axis is represented the Mahalanobis distance of one sample against the previously seen 120 samples. The last seen sample is at abscissae 120, which explains why the Mahalanobis distance is close to zero on that point. On the z-axis, represented by different tones in the colour blue, is represented the history of this computation over 50 samples. In other words, the range of curves represents the evolution of the Mahalanobis distance of a collection of samples over time. For repetitive motions, we observe another minimum, marked here with a vertical line at abscissae 30, which marks the period of the motion. This collection of curves resemble a cycloid, the trajectory a point follows when attached to a wheel. This kind of structure is not observed when random motions are submitted to the same treatment. The structure in this buffer seem to be salient enough to perform a visual classification. These are thus chosen as the features for our discriminative model.

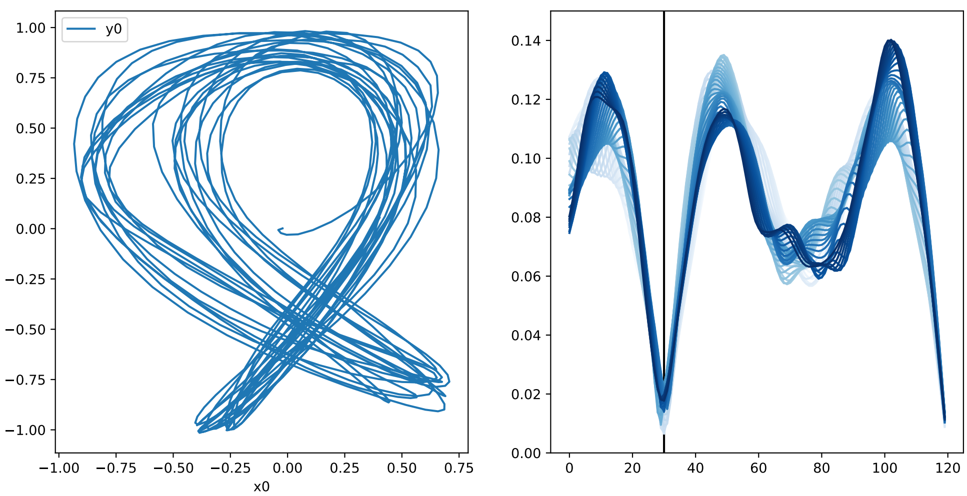

Figure 5.10: Periodic motion and associated features on the left and right subplots, respectively. The period of the motion is indicated by a vertical line on abscissae 30.

Now that the features are defined, the problem of classifying them can be addressed. The data we described has both a space and time component. We chose to use a neural network based

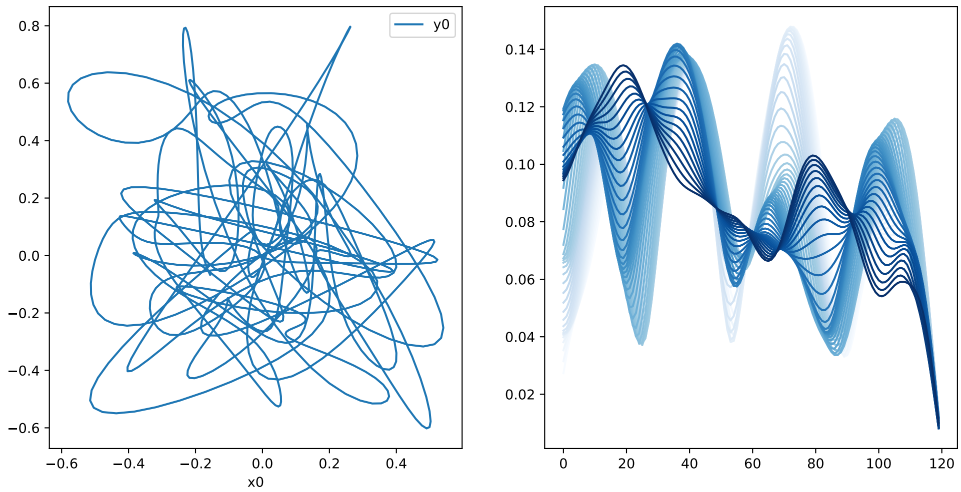

Figure 5.11: Random walk and associated features, on the left and right subplots, respectively. The features for a non-periodic motion do not exhibit a strong stability along the colour dimension at any point along the abscissae.

on 1D convolution, as their performance as been proven superior for this task [117]. The architecture of the model is detailed in Table 5.5 and has been adapted from recommended code samples available from the library Keras [72] used for the computation.

Layer	Output Shape	Setting	Param #
input	5, 20	n.a.	0
conv_1d	3, 16	relu	5776
conv_1d	1, 16	relu	784
dropout	16	50%	0
flatten	16	50%	0
dense	1	sigmoid	17

Table 5.5: Neural network architecture with a total of 6577 parameters.

We generated a dataset from sampling our synthetic models of motions. We created one million samples for the training set and two hundred thousands samples for the test set. The performance given by the confusion matrix and the operating curve are plotted on Figure 5.12.

An additional test was carried out to investigate the performance of our model with different levels of noise. An additional set of synthetic samples was generated, with noise level spanning 0% to 100% of the amplitude of the oscillator. The results from this sensitivity analysis are presented on Figure 5.13. It appears that the model performance is reasonably stable across the set of noise levels. There is however one outlier. When the noise level is 0%, that is the motions are perfectly periodic, the performance drops. This corner case might not be

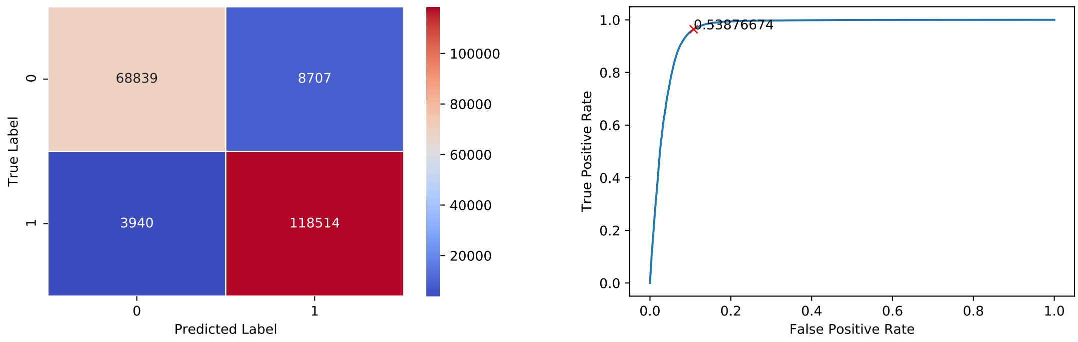

Figure 5.12: Performance of the classification between samples generated from models of random and noisy cyclical motions.

seen in data generated by participants, since it is unlikely that human subjects can perform physical motions with perfect repetitions.

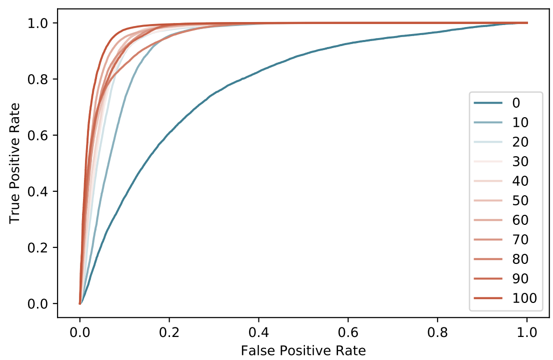

Figure 5.13: Sensitivity of discriminative model to different levels of noise, spanning 0% to 100% of the amplitude of the oscillator generating the cyclical motion.

5.6.3 Segmentation

Finally, it is also interesting to segment or count the repetitions within a motions that would have been classified as repetitive by the model developed in the previous section. For this purpose, a simple algorithm was designed which exploits the observation that a stable minimum appears in the 2-dimensional buffer that stores the Mahalanobis distance. A seed for the motion is stored as the first sample that is classified as repetitive. The Mahalanobis distance between the history and this seed is computed, and in the presence of a global minima a repetition is declared completed. While the motion is classified as repetitive, every sample that completes a period is stored. The output of the algorithm is a list of time stamps describing each completed cycles. The pseudo code for this algorithm is detailed below:

Algorithm 2 Segment repetitions

```
in_repetition ← false
while new sample do
    features ← transform(sample)
    repet ← model(features)
    is_repetitive ← repet > threshold
    history.pop_last()
    history.append(sample)
    if not is_repetitive and in_repetition then
        this is the last sample for the repetition
    end if
    if in_repetition then
        this is a new sample for the current repetition
        if D_M(seed, history) is global minima then
            this is a complete period, save the sample
        end if
    end if
    if is_repetitive and not in_repetition then
        this is the first sample of a new repetition, save the seed
        seed ← sample
    end if
end while
```

This last section completes the design of the system that is needed for the recognition and segmentation of repetitive motions in joint user-sensor spaces. With such a classifier and the segmentation algorithm, a continuous stream of samples from the joint user-sensor space can be analysed in real-time, its repetitive nature assessed and the number of repetitions within a repetitive section can be counted. As a result, an experiment can be undertaken in which participants are rewarded for the repetitive nature of their motions.

5.7 RTR Experiment

We designed an experiment with the following questions in mind:

- What kind of motions do users produce when rewarded for repetitiveness?
- What volume do repetitive motions occupy and how does it compare with the volume observed in the RTO experiment?

5.7.1 Apparatus

A desk providing an interaction area of $78cm$ by $78cm$ was overlooked by an Intel Realsense camera, which through custom software provided the optical tracking of a marker that participants were wearing in between their index and middle finger and attached by a stretchable strap. This setup is similar to the one employed for the RTO experiment, see 5.5. An office chair was facing the camera and the interaction area.

The computer running the tracking software was also responsible for running the signal processing that transformed readings from the tracker into 6-dimensional vectors from the joint user-sensor space. From the samples, our model was classifying in real-time the repetitive nature of the currently observed motion. The audio feedback responsible for conveying the reward to the participants was designed as such: a background sine wave at $90Hz$ would indicate the default state (not repetitive), and would be pitched up to $180Hz$ when the motions produced by the participant were deemed repetitive by the output of the classifier. Once in repetition, the segmentation algorithm would count the number of current repetition and on the 10th cycles a high-pitched feedback would be produced at $270Hz$ with a $200ms$ decay, indicating the completion of the task to the participants.

In addition, the state change of the audio feedback from neutral to repetitive was delayed with 2 full cycles. This prevented participants from being subject to the initial fluctuations in the output of the classifier at the beginning of a repetitive cycle and from being disturbed by the audio feedback before having locked themselves into a stable motion pattern.

5.7.2 Task

The task for participants was to produce a cyclical motion of at least ten full period. The completion of the task was conveyed through the audio feedback.

5.7.3 Procedure

Participants were seated by the desk and equipped with a marker. They were not given any postural recommendations. Participants were explained the task. They were instructed that the audio feedback would reward repetitive or cyclical motions of their hand, and that the position, velocity and acceleration would be taken into account as part of the repetitive nature of their motions. They were then informed about how the audio feedback was constructed with the differences between the three pitches. Finally, they were also instructed to emphasise originality across repetitive motions.

5.7.4 Design

We ran the task with ten repetitions. As it was impossible to fail or abort the task, the number of repetitions was chosen to keep the experiment under ten minutes, provided one task was completed per minute.

5.7.5 Participants

We recruited 6 participants, including 2 female participants, with a mean age of 35.5 and standard deviation of 11.2 (min=23, max=55). All participants were right handed. The study was approved by the University of Glasgow ethics committee. No participants presented any mental or physical disabilities. These were the same participants that took part in the RTO experiment. They were as such familiar with the concept of reward through audio feedback.

5.7.6 Results

The data collected through the experiment was composed of the samples from the joint user-sensor space, time stamped and evenly sampled in time at $30Hz$. For each sample, the output of the classifier was also recorded.

The experiment lasted for $4.8min$ on average with a standard deviation of $2.4min$, a minimum at $2.7min$ and a maximum at $9.3min$, indicating that all participant did manage to produce ten cyclical motions of ten repetitions each, as measured by our algorithms. We also computed the percentage of time the participant did produce motions that were deemed repetitive. On average, participant motions were deemed repetitive 68% of the time with a standard deviation of 14%, a minimum at 41% and a maximum at 84%. The participant that took the longest time to complete the experiment was not the participant that had the lowest fraction of time with repetitive motions. Instead, it is the difficulty to complete ten consecutive repetitions of a single motion that rather explains the time taken for the experiment.

All the cyclical motions with their full ten repetitions are plotted on Figure 5.14, with one participant per columns. The last row on the figure shows the samples that were excluded from this segmentation process. For clarity, we did not include the unit on the individual plots, but they are all centred on the origin, plotted with equal proportion on both axis and the grid is sampled every $20cm$. All motions thus fit within a square of $80cm$ by $60cm$ which is similar to the interaction area that was proposed to the participants. Regarding the shapes produced by the participants, linear motions produced by a back and forth between two inflection points are the most common with 20 occurrences. All participants produced that shape, with the extreme case of the participant plotted on the last column who did produce it exclusively. Figure of 8 follows in terms of frequency of appearance with 9 occurrences (clustered in two groups, one having sharp corners) and their presence is observed in all participant but two. Six motions present a very small amplitude and should be considered as false positive from our classifier. Circles are also present with four occurrences, which could also be clustered together with one ellipse. The rest of the shapes are triangles, arrows, squares, figures of 8 with more twists and complex shapes which result of the composition of these elementary patterns.

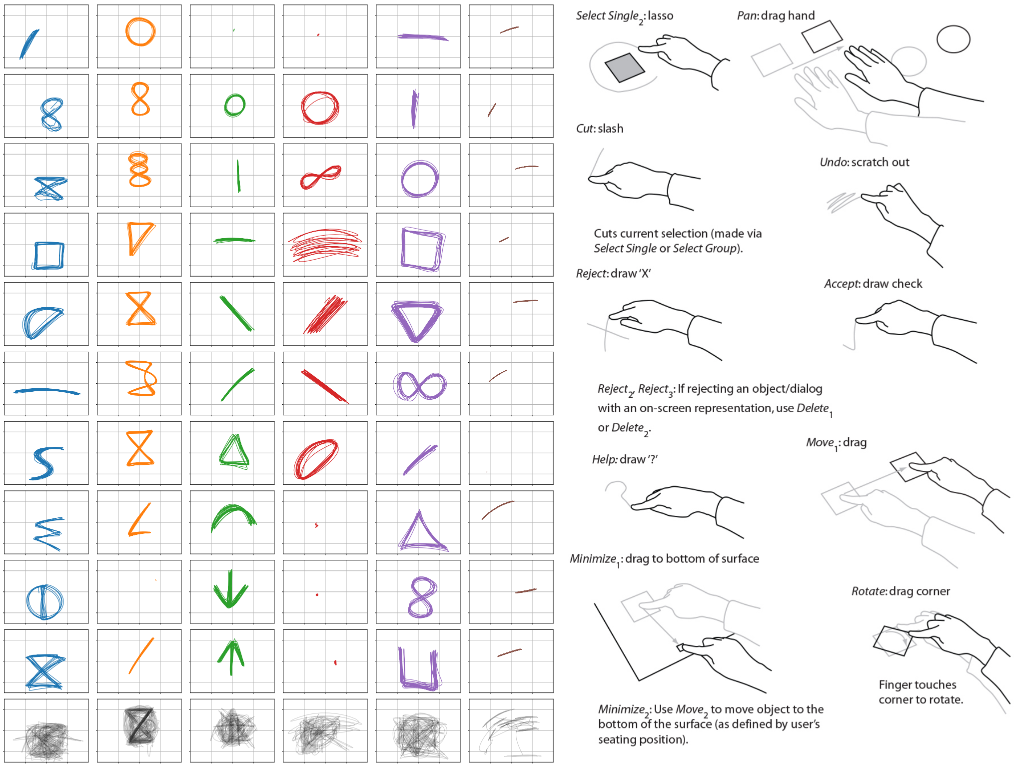

Figure 5.14: Dataset of cyclical motions collected through the RTR process plotted in X0 and Y0, one columns per participants. For clarity, tick labels are removed, squares have dimension $20cm$ by $20cm$ and subplots' centres are at the origin. The last row displays the data points that were considered as not part of the task during the experiment (left). Chosen extract of the motions produced by the elicitation process for surface computing, reproduced and adapted from [Wobbrock - 2009] (right).

For comparison purpose, and to draw similarities between the process of elicitation in HCI, the output of an elicitation study for surface computing by Wobbrock et al. [37] is reproduced by the side of the motions that were produced by RTR, see Figure 5.14. The shapes selected present some similarities, such as geometrical pattern (circle and cross) or linear and circular motions. One main difference for the shapes omitted lies in the presence of bi-manual gestures, which our experiment did not support or composite touching gestures which required users to lift their fingers. These missing motions could however be also integrated in the process proposed by RTR.

Finally, the volume of the joint user-sensor space occupied by these repetitions was computed and compared to the volume of the joint user-sensor space captured in the RTO experiment, see Table 5.6. Overall, the mean value of the logarithm of the volume was smaller for the RTR experiment than for the RTO experiment. The variation across participants was also smaller for the RTR experiment than for the RTO experiment. It was expected that the repetitive motions consist in a subset of all possible motions, hence the smaller volume.

	$log(V_{RTR})$	$log(V_{RTO})$
mean	-11.6	-9.53
std	2.5	3.83
min	-14.2	-14.27
max	-8.4	-5.88

Table 5.6: Volume for the joint user-sensor spaces measured during the RTO and RTR experiments.

Discussion and Future Directions

To identify repeatability in motions, the concept of cyclical motions was used and a model has been built to produce synthetic samples of such category. The basis for this decision comes from a study from Guiard [118] in which it is argued that motions are not a concatenation of movement primitives but rather should be described as oscillations. And that despite the prevalence of pointing motions in HCI, which are usually described as half-cycle, fully cyclical motions are potentially the more general building blocks of motions. They have indeed been successfully modelled as such [58]. However, considering the results from the comparison of motions elicited by RTR and an elicitation study for surface computing [37], it is clear that more diversity in terms of motions is required to cover a bigger proportion of what can be produced in elicitation studies. This could, for instance, include motions which have a pause in between two full-cycles. These were not present in the dataset but could be created.

Also, it would be interesting to see how the properties of the classification model influences the stability of the motions produced by participants. In the RTR experiment, the threshold for the operating point of the classifier was kept constant but it is possible to chose any value between 0 and 1. The same idea applies to the type of data we provided the model for its training. We chose a fixed value of for the amplitude of the Perlin noise at 30% of the maximum amplitude of the cyclical motions. By increasing or reducing this value, it could be possible to generate different levels of stability in the cyclical motions. For a level of 0%, the repetitions are identical, indicating a perfect stability across time of the motions, while for bigger value, the noise level tends to overtake the signal represented by the cyclical motions. At the extreme, cyclical motions are undistinguishable in the noise and their properties should be similar to those of random motions.

The audio feedback that was devised in the original experiment was conveying only the repetitiveness of motions back to the user. In the RTO experiment, a notion of originality was automatically fed back to the users. This allowed for a search process to take place in which users would explore the originality in the motions they could produce. In the RTR experiment, this feedback was lacking. Therefore, we instructed the participants to produce repetitions that they deemed original. Including a feedback on the originality of repetitions as compared to previously recorded ones consists an important extension to the technique we presented. However, more complex models of motions are needed here to provide such a measure.

Finally, the techniques presented here are not specific to upper limb gestural interactions. They could be applied to any other joint user-sensor space such as the ones describing interactions with mobile devices.

5.8 Conclusion

This chapter has defined joint user-sensor spaces and different measures that can be used for their quantification. Drawing from the method of elicitation used in HCI, a process for the exploration of these spaces was revisited and applied to the specific case of upper limb gestural interactions on planar surfaces. The volume of such spaces was compared for the different cases of the interactions we have proposed in the previous chapters and displayed that the gesture typing interaction occupied the smallest interaction volume of all investigated interactions. A weak correlation was found between interaction volumes and information throughput which introduced the need for measuring the repeatability of motions produced in joint user-sensor spaces. A discriminative model was trained on a synthetic dataset which included random motions and cyclical noisy motions generated from a low-pass filtered random walk, and the IDFT of a random spectrum with additive Perlin noise, respectively. A user study captured a collections of motions participants deemed repetitive, which was compared to the output of an elicitation study aimed at surface computing gestures.

Chapter 6

Conclusion

There are many potent reasons for the introduction of new gestural interactions. Opportunities afforded by the use of new sensors, the inclusion of novel body parts, or using a task in previously unexplored scenario are all valid reason for the exploration of novel gestural interactions. For instance, this thesis proposed to port text-input to optically tracked surface, potentially making use of their low reliance of fine finger motor control and opportunities for large interaction area. This thesis also proposed to repurpose an arm reach rehabilitation exercise for digital game control and showed that an Arcade game such as *Pac-Man* can be successfully played through an upper limb gestural interaction.

However, after an initial sketch or design phase, the introduction of new interactions results in challenges with regards to their creation, optimisation or understanding. New sensors often requires dedicated models for processing raw input data stream into variables users can control, the many parameters and unexpected effects can lead to challenging optimisation problems and user studies often reveal unexpected or unforeseen behaviours.

In this thesis, we have proposed to use a computational method to model these challenges. We used concepts directly taken from a computing approach to frame gestural interaction as computing problems, which granted access to tools usually employed in such scenarios. For instance, framing the detection of touch contact between fingers and tabletops in RGB-D image as a classification task naturally led to the use of supervised learning techniques. Using digital games for physical rehabilitation was decomposed as a dual optimisation process between in-game performance and user movements, for which design configuration can be found provided an objective function with low sampling latency is available. Finally, the task of eliciting motions was redefined as a search process in which users where guided through by an audio feedback.

6.1 Summary of Contributions

This thesis made a list of contributions, see section 1.2. Three types of contributions have been made including new gestural interactions of the upper limb on planar surfaces, models for their creation using machine learning techniques and results from tests and users studies carried out with human participants.

New interactions have been proposed. In particular, Chapter 3 introduced the task of text input through optically tracked surfaces using the technique of gesture typing. This was executed in an indirect absolute manner with varying scale and control/display gain. Chapter 4 proposed to perform the rehabilitation of arm reach through a gamified interaction. This used off-the-shelf digital games, for instance *Pac-Man*, thanks to the creation of a special input control modality which purpose was to serve as a interface to the game control mechanism while requiring the motions needed for rehabilitation. Finally, Chapter 5 advanced the idea of automated elicitation of motions specific to the upper limb on planar surfaces where both variability and repeatability of motions were sough after. This was relying on an audio feedback produced according to the real-time observed properties of motions.

Different models have been developed to enable these interactions. Discriminative models have played a central role, with the use of neural networks in Chapter 3 and 5. These were trained from datasets either collected from human participants or created synthetically from other models supposed to simulate behaviours that were expected to be seen with human participants. The reliance on datasets of interaction for HCI is a crucial step towards more reproducibility. Probabilistic models of behaviours were also used, for instance in Chapter 4, where the reference gameplay from participants was summarised as the product of low-level variables' distributions. These distributions were fitted to human participants data and were employed to assess the likelihood of newly observed behaviours to be generated from the same models.

To validate our design choices and answer our research questions, user studies and tests in hospital settings have been conducted. In Chapter 3, the influence of scale on the performance of text-input in an indirect absolute gesture typing task was measured. This revealed that this interaction was slower than a control interaction on a tablet, and that the influence of scale was measurable only on the error rate but not on the input rate. In Chapter 4, the impact of a new control modality on user in-game performance was measured and was shown to be negative according to expectations. We demonstrated that instrumenting the time-rate of the game could counteract this effect. The influence of different locations of the virtual control positions were also recorded and showed a positive, yet non-linear, effect on participant behaviour who overshot consistently their targets. Finally, in Chapter 5, the variability of motions for upper limb interaction was recorded across five participants, shedding some light on the structure of their physical ability in terms of speed and acceleration of their end-effectors. The maps that were produced constitute a potential tool for OTs in the quantitative measure of human performance. The repeatability of motions was also recorded and showed some similarities with motions elicited through conventional elicitation techniques.

In other words, the contributions in Chapter 3 add to the work of Xiao et al. [32] by introducing a new interaction to optically tracked surfaces, and by introducing machine learning techniques to the processing needed to extract information from RGB-D images. Also, it adds another data point to studies related to effect of scales on steering tasks [52], input sizes on gesture typing [67], indirect gesture typing [66] or comparison between indirect and direct input techniques [119]. The contributions in Chapter 4 are meant to extend the frameworks proposed by Walther-Franks et al. [120] and Ketcheson et al. [88] by viewing the use of off-the-shelf games as an optimisation process. This chapter clearly identifies the role of the input control modality and provides a solution to the problem of balancing its effect on user performance, which has been flagged as a long-standing issue [100]. Lastly, the contributions in Chapter 5 offer an extension to the work of Williamson et al. [107] by tackling the problem of classifying motion repetitions in continuous sensor streams. Also, a clear connection with elicitation studies [36] has been established.

This thesis proposed to explore how a computational approach to gestural interactions can address the challenges their creation, optimisation or understanding present. The list of requirements for a computational approach is reproduced here for clarity:

1. an explicit mathematical model of user-system behaviour;
2. a way of updating that model with observed data from users;
3. an algorithmic element that, using this model, can directly synthesise or adapt the design;
4. a way of automating and instrumenting the modelling and design process;
5. the ability to simulate or synthesise elements of the expected user-system behaviour.

We can see that all but one of the elements of this list were used. The model of touch contact between fingers and tabletops from Chapter 3 implements the first item, while the model of reference gameplay from Chapter 4 fits the second item. The use of the digital game's time rate for user experience optimisation in Chapter 4 is linked to the third item and the models of synthetic motions such as the random walk and oscillators from Chapter 5 are relevant to the last item. The fourth item, seeking to automate the design process, was not approached. The design process was instead conducted through discussions and design workshops with OTs.

6.2 Limitations

Despite the previous list of contributions made by this thesis, there are a number of limitations to this work.

Role of Motor Impaired Users in Experimental Studies

One of the main limitations lies in the relative lack of participants who have sustained a SCI in experimental studies. This shows throughout the whole thesis but is especially prevalent in Chapter 3, in which no physically impaired participants had the opportunity to test the system affording gesture typing through optically tracked surfaces. During a discussion with OTs, the feedback they provided hinted that the design of the interaction itself might not be well adapted to users with SCI. These were accustomed to using their mobile devices and used various mitigation strategies to interact with common touchscreens available on smartphones, as pointed out in research studies such as the one carried out by Anthony et al. [18]. This highlights the importance of participatory design in the establishment of a new interaction. In Chapter 4, some participants with SCI were involved and provided very valuable qualitative feedback validating the decisions made until then. It was however not possible to quantify their interactions with the digital game and measure their hand movements on the tabletop. This would have provided some data for measuring the potential of this interaction for upper limb reach rehabilitation. The last chapter would also have benefited from participation of users with motor impairment, especially when it comes to the data visualised as a reachable map or maximum velocity and acceleration maps. Some differences were already observed between participants of the user study, but comparisons across different user groups are usually very informative, such as the one produced by Findlater et al. [22]. There exists thus some potential for extending the present work with further studies involving participants with motor impairments. The interactions with OTs and patients in the QEU hospital has however been immensely valuable, greatly helped finding innovative ideas and topic for research that could benefit a larger population. This echoes the findings from research focused on OTs [121].

Design of Experimental Studies

The user experiments presented in this thesis did recruit a relative small numbers of participants with 12 maximum users enrolled. For Chapter 3, this number was enough to uncover some significant results. However, the conclusions drawn from for the experiment in Chapter 5 are limited in their applicability to a more general population.

Chapter 3

The main limitation in this chapter consists in the lack of comparison of the performance of the touch classification with other available systems or algorithms. Even though these were published roughly at the same time as my research was carried out, the release of datasets is unfortunately not yet a common practise in the field of HCI, preventing easily repeatable experiments. To add to the research community, the source code and the dataset for the detection of touch contact has been made publicly available.

Chapter 4

This chapter has been setting the tools for an optimisation of the design to take place, but the user study has not been carried out. As a result, only the measure of performance with low-latency has been considered as a contribution. A live optimisation of play sessions with players having sustained a SCI would provided valuable information as to how finding suitable configurations of design parameters can be carried out, especially in association with OTs. Another shortcoming is the lack of user engagement modelling. This is usually done through a lengthy qualitative procedure and was only partially done in this work through interviews with the participants of the user study.

Chapter 5

The performance of the models used for detecting repetitions is highly dependent on the data used for their training. The comparison with the results of the elicitation study for surface computing [37] shows that better and more generic models are required if this technique is meant to complement elicitation studies. This could be solved with a richer synthetic dataset for example or by using example of motions from real participants. Also, the question of the definition of noise in the synthetic models remains opened. The chosen technique to simulate human motor limitations was to simply add Perlin noise to the amplitude of synthetic motion trajectories. Given the results from the user study and preliminary tests with other input modalities, such as a smartphone, the discriminative model appears to fulfil its intended purpose but it is still unclear whether this noise model is realistic enough to elicit very high quality repetitions for example.

6.3 Future Work

Each of the research activity conducted in Chapter 3, 4 and 5 do open several avenues for research.

6.3.1 New Avenues for Optically Tracked Surfaces

Given that continuous interaction, such as gesture typing, appear to be suitable for optically tracked surfaces, others ecologically valid task should also be investigated. It is not unrealistic to imagine that such interactions could complement regular interactions with mobile devices. These would benefit from the added real estate provide by such systems, or the possibility to instrument the texture [122], the relief and topology [123] of the interaction surface. Another complementary avenue for research is to leverage the ability of tracking the user pointer when it is not in contact with the surface, opening a lot of potential for the association of touch and mid-air input. More research is however needed before such interactions are commonplace. In particular, understanding the interplay between the user and the visual hover feedback provided on the screen is of prime importance, since it was observed that a non-negligible offset on touch down was produced by the user motion.

6.3.2 New Avenues for Digitally-aided Physical Rehabilitation

Drawing from the limitations of this chapter, a potential research focus would consist in a more thorough study of the parameter optimisation. The function that represents the effect of SPREAD and T_RATE on NLL seems to be monotonic and not very noisy. Different optimisation technique could be employed for finding a suitable solution, but in practise these remains to be seen. Beyond that objective, an investigation of the user engagement in relation to low-level variables would be very interesting. In the current work, the user engagement was assumed to correlates with the user performance. This model has some limitations and understanding what in particular do players enjoy in their playful interaction could lead to better design or smarter decision in the making of alternative input control modalities. This is also related to solving the task of matching games with rehabilitation exercises. This challenge has not been the focus on the present research, and a solution was found through participatory design sessions involving HCI researchers and OTs. But more complex or widely different exercising than upper limb reach rehabilitation are also in need of increased user motivation and the mapping between control and motions might not be as straightforward as in the present work. Here, informing the design aided with a computational approach seem to be a promising way forward.

6.3.3 New Avenues for an Automatic Elicitation Process

The need for finding motions that are suitable for gestural interaction does not yet have an formalised approach when it comes to modelling user capabilities. The present approach does provides with a way forward and it would be important to understand how this new data can be fed into a design process. For instance, the non-uniformity in distributions of maximum velocities and accelerations reached during the experiment could be used for selecting easily performable, yet recognisable, gestures. Also, variability and repeatability were exclusively targeted in this research, but other properties could be envisioned. The invention time has been proposed as an interesting quantifiable property [107]. Understanding or defining this in relation to the concept of a motion to be natural is a promising undertaking. Finally, the present research was limited to upper limb motions on planar surfaces, but interactions with smartphones or smartwatches, making use of IMUs would be very opportune to investigate.

6.4 Summary and Conclusion

Novel gestural interactions provide for valuable new experiences. Upper limb interactions on planar surfaces, despite their defining limitations, have be shown to afford a variety of scenarios, such as text-input or Arcade gaming. After an initial design phase, the creation, optimisation and understanding of new interactions bring challenges a computational approach can address. It provides frameworks with strong descriptive capabilities and tools for solving computing problems. Classification of images, optimisation and search process are among the techniques employed here to support this assertion.

Some aspects of computational interaction design were not approached in this thesis, in particular looking for ways of “automating and instrumenting the modelling and design process.” This goes beyond what has been proposed in this work, but appears to be a natural extension to the work on physical rehabilitation through games, and poses the important question of the relationship between the task of design and computational design.

Bibliography

[1] A. D. Wilson, "Using a depth camera as a touch sensor," in *Proc. ITS 2010*. New York, New York, USA: ACM Press, nov 2010, pp. 69–72. [Online]. Available: http://dl.acm.org/citation.cfm?id=1936665

[2] F. Anderson, T. Grossman, J. Matejka, and G. Fitzmaurice, "YouMove," in *Proc. 26th Annu. ACM Symp. User interface Softw. Technol. - UIST '13*. New York, New York, USA: ACM Press, 2013, pp. 311–320. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2501988.2502045

[3] J. Rekimoto, "Tilting operations for small screen interfaces," in *Proc. 9th Annu. ACM Symp. User interface Softw. Technol. - UIST '96*. New York, New York, USA: ACM Press, 1996, pp. 167–168. [Online]. Available: https://dl.acm.org/citation.cfm?doid=237091.237115

[4] G. Bailly, J. M?ller, M. Rohs, D. Wigdor, and S. Kratz, "ShoeSense," in *Proc. 2012 ACM Annu. Conf. Hum. Factors Comput. Syst. - CHI '12*. New York, New York, USA: ACM Press, 2012, p. 1239. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2207676.2208576

[5] P.-O. Kristensson and S. Zhai, "SHARK 2," in *Proc. 17th Annu. ACM Symp. User interface Softw. Technol. - UIST '04*. New York, New York, USA: ACM Press, oct 2004, p. 43. [Online]. Available: https://dl.acm.org/citation.cfm?doid=1029632.1029640

[6] P. Vogiatzidakis and P. Koutsabasis, "Gesture Elicitation Studies for Mid-Air Interaction: A Review," *Multimodal Technol. Interact.*, vol. 2, no. 4, p. 65, sep 2018. [Online]. Available: http://www.mdpi.com/2414-4088/2/4/65

[7] A. Toney and B. Thomas, "Considering Reach in Tangible and Table Top Design," in *First IEEE Int. Work. Horiz. Interact. Human-Computer Syst. (TABLETOP '06)*, vol. 2006. IEEE, 2006, pp. 57–58. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1579192

[8] J. Napier, J. R. Napier, and R. H. Tuttle, *Hands*. Princeton University Press, 1993.

[9] A. Crossan, J. Williamson, S. Brewster, and R. Murray-Smith, "Wrist rotation for interaction in mobile contexts," in *Proc. 10th Int. Conf. Hum. Comput. Interact. with Mob. devices Serv. - MobileHCI '08*. New York, New York, USA: ACM Press, 2008, p. 435. [Online]. Available: https://dl.acm.org/citation.cfm?doid=1409240.1409307

[10] H. Pohl and R. Murray-Smith, "Focused and casual interactions," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst. - CHI '13*. New York, New York, USA: ACM Press, apr 2013, p. 2223. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2470654.2481307

[11] R. Müller-Putz, G.R., Ofner, P., Schwarz, A., Pereira, J., Luzhnica, G., di Sciascio, C., Veas, E., Stein, S., Williamson, J., Murray-Smith, R., Escolano, C., Montesano, L., Hessing, B., Schneiders, M., and Rupp, "Moregrasp: Restoration of Upper Limb Function in Individuals with High Spinal Cord Injury by Multimodal Neuroprostheses for Interaction in Daily Activities," in *7th Graz Brain-Computer Interface Conf.*, Graz, Austria, 2017, pp. 338–343.

[12] K. D. Anderson, "Targeting Recovery: Priorities of the Spinal Cord-Injured Population," *J. Neurotrauma*, vol. 21, no. 10, pp. 1371–1383, oct 2004. [Online]. Available: http://www.liebertonline.com/doi/abs/10.1089/neu.2004.21.1371

[13] G. J. Snoek, M. J. IJzerman, H. J. Hermens, D. Maxwell, and F. Biering-Sorensen, "Survey of the needs of patients with spinal cord injury: impact and priority for improvement in hand function in tetraplegics," *Spinal Cord*, vol. 42, no. 9, pp. 526–532, jun 2004. [Online]. Available: http://www.nature.com/doifinder/10.1038/sj.sc.3101638

[14] F. M. M. Jr, M. B. Bracken, G. Creasey, J. F. D. Jr, W. H. Donovan, T. B. Ducker, S. L. Garber, R. J. Marino, S. L. Stover, C. H. Tator, R. L. Waters, J. E. Wilberger, and W. Young, "International Standards for Neurological and Functional Classification of Spinal Cord Injury," *Spinal Cord*, vol. 35, no. 5, pp. 266–274, 1997. [Online]. Available: http://www.nature.com/doifinder/10.1038/sj.sc.3100432

[15] G. Kayalioglu, *The Vertebral Column and Spinal Meninges*. Elsevier Ltd, 2009. [Online]. Available: http://dx.doi.org/10.1016/B978-0-12-374247-6.50007-9

[16] K. Church and B. Smyth, "Understanding the intent behind mobile information needs," in *Proceedingsc 13th Int. Conf. Intell. user interfaces - IUI '09*. New York, New York, USA: ACM Press, 2008, p. 247. [Online]. Available: https://dl.acm.org/citation.cfm?doid=1502650.1502686

[17] T. Sohn, K. A. Li, W. G. Griswold, and J. D. Hollan, "A diary study of mobile information needs," *Proceeding twentysixth Annu. CHI Conf. Hum. factors Comput. Syst. CHI 08*, pp. 1–10, 2008. [Online]. Available: https://dl.acm.org/citation.cfm?doid=1357054.1357125

[18] L. Anthony, Y. Kim, and L. Findlater, "Analyzing user-generated youtube videos to understand touchscreen use by people with motor impairments," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst. - CHI '13*. New York, New York, USA: ACM Press, 2013, p. 1223. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2470654.2466158

[19] M. Naftali and L. Findlater, "Accessibility in context: understanding the truly mobile experience of smartphone users with motor impairments," *Proc. 12th Int. ACM SIGACCESS Conf. Comput. Access. - ASSETS '14*, pp. 209–216, 2014. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2661334.2661372

[20] K. Montague, H. Nicolau, and V. L. Hanson, "Motor-impaired touchscreen interactions in the wild," in *Proc. 16th Int. ACM SIGACCESS Conf. Comput. Access. - ASSETS '14*. New York, New York, USA: ACM Press, 2014, pp. 123–130. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2661334.2661362

[21] T. Guerreiro, H. Nicolau, J. Jorge, and D. Gonçalves, "Towards accessible touch interfaces," in *Proc. 12th Int. ACM SIGACCESS Conf. Comput. Access. - ASSETS '10*. New York, New York, USA: ACM Press, 2010, p. 19. [Online]. Available: https://dl.acm.org/citation.cfm?doid=1878803.1878809

[22] L. Findlater, K. Moffatt, J. E. Froehlich, M. Malu, and J. Zhang, "Comparing Touchscreen and Mouse Input Performance by People With and Without Upper Body Motor Impairments," in *Proc. 2017 CHI Conf. Hum. Factors Comput. Syst. - CHI '17*. New York, New York, USA: ACM Press, 2017, pp. 6056–6061. [Online]. Available: http://dl.acm.org/citation.cfm?doid=3025453.3025603

[23] A. M. Piper, R. Campbell, and J. D. Hollan, "Exploring the accessibility and appeal of surface computing for older adult health care support," in *Proc. 28th Int. Conf. Hum. factors Comput. Syst. - CHI '10*. New York, New York, USA: ACM Press, apr 2010, p. 907. [Online]. Available: http://dl.acm.org/citation.cfm?id=1753326.1753461

[24] M. Augstein, T. Neumayr, R. Ruckser-Scherb, I. Karlhuber, and J. Altmann, "The fun.tast.tisch. project– A Novel Approach to Neuro-Rehabilitation Using an Interactive Multiuser Multitouch Tabletop," *Proc. 2013 ACM Int. Conf. Interact. tabletops surfaces - ITS '13*, pp. 81–90, 2013. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2512349.2512808

[25] M. Augstein, T. Neumayr, and I. Schacherl-Hofer, "The Usability of a Tabletop Application for Neuro-Rehabilitation from Therapists' Point of View," in *Proc. Ninth ACM Int. Conf. Interact. Tabletops Surfaces - ITS '14*. New York, New York, USA: ACM Press, 2014, pp. 239–248. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2669485.2669516

[26] G. Pullin, *Design meets disability*. MIT Press, 2009.

[27] K. Shinohara and J. O. Wobbrock, "In the shadow of misperception," in *Proc. 2011 Annu. Conf. Hum. factors Comput. Syst. - CHI '11*. New York, New York, USA: ACM Press, 2011, p. 705. [Online]. Available: http://dl.acm.org/citation.cfm?doid=1978942.1979044

[28] P. Wellner, "Interacting with paper on the DigitalDesk," *Commun. ACM*, vol. 36, no. 7, pp. 87–96, jul 1993. [Online]. Available: http://dl.acm.org/citation.cfm?id=159544.159630

[29] P. Dietz and D. Leigh, "DiamondTouch," in *Proc. 14th Annu. ACM Symp. User interface Softw. Technol. - UIST '01*. New York, New York, USA: ACM Press, 2001, p. 219. [Online]. Available: https://dl.acm.org/citation.cfm?doid=502348.502389

[30] O. Hilliges, A. Butz, S. Izadi, A. D. Wilson, and A. D.Wilson, "Interaction on the Tabletop: Bringing the Physical to the Digital," in *Tabletops – Horiz. Interact. Displays*. Springer, London, 2010, pp. 189–221. [Online]. Available: http://link.springer.com/10.1007/978-1-84996-113-4_9

[31] C. Harrison, H. Benko, and A. D. Wilson, "OmniTouch," in *Proc. 24th Annu. ACM Symp. User interface Softw. Technol. - UIST '11*. New York, New York, USA: ACM Press, oct 2011, p. 441. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2047196.2047255

[32] R. Xiao, S. Hudson, and C. Harrison, "DIRECT: Making Touch Tracking on Ordinary Surfaces Practical with Hybrid Depth-Infrared Sensing," *Proc. 2016 ACM Interact. Surfaces Spaces - ISS '16*, pp. 85–94, 2016. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2992154.2992173

[33] R. Xiao, J. Schwarz, N. Throm, A. D. Wilson, and H. Benko, "MRTouch: Adding touch input to head-mounted mixed reality," *IEEE Trans. Vis. Comput. Graph.*, vol. 24, no. 4, pp. 1653–1660, 2018.

[34] M. E. Mott, R.-D. Vatavu, S. K. Kane, and J. O. Wobbrock, "Smart Touch," in *Proc. 2016 CHI Conf. Hum. Factors Comput. Syst. - CHI '16*. New

York, New York, USA: ACM Press, 2016, pp. 1934–1946. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2858036.2858390

[35] A. Kendon, *Gesture : visible action as utterance*. Cambridge University Press, 2004.

[36] J. O. Wobbrock, H. H. Aung, B. Rothrock, and B. A. Myers, "Maximizing the guessability of symbolic input," in *CHI '05 Ext. Abstr. Hum. factors Comput. Syst. - CHI '05*. New York, New York, USA: ACM Press, 2005, p. 1869. [Online]. Available: https://dl.acm.org/citation.cfm?doid=1056808.1057043

[37] J. O. Wobbrock, M. R. Morris, and A. D. Wilson, "User-defined gestures for surface computing," in *Proc. 27th Int. Conf. Hum. factors Comput. Syst. - CHI 09*. New York, New York, USA: ACM Press, 2009, p. 1083. [Online]. Available: http://dl.acm.org/citation.cfm?doid=1518701.1518866

[38] J. Ruiz, Y. Li, and E. Lank, "User-defined motion gestures for mobile interaction," in *Proc. 2011 Annu. Conf. Hum. factors Comput. Syst. - CHI '11*. New York, New York, USA: ACM Press, 2011, p. 197. [Online]. Available: http://dl.acm.org/citation.cfm?doid=1978942.1978971

[39] V. Lantz and R. Murray-Smith, "Rhythmic interaction with a mobile device," in *Proc. third Nord. Conf. Human-computer Interact. - Nord. '04*. New York, New York, USA: ACM Press, 2004, pp. 97–100. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1028014.1028029

[40] K. Wolf, S. Mayer, and S. Meyer, "Microgesture detection for remote interaction with mobile devices," in *Proc. 18th Int. Conf. Human-Computer Interact. with Mob. Devices Serv. Adjun. - MobileHCI '16*. New York, New York, USA: ACM Press, 2016, pp. 783–790. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2957265.2961865

[41] H. Pohl, M. Krause, and M. Rohs, "One-button recognizer," in *Proc. 2015 ACM Int. Jt. Conf. Pervasive Ubiquitous Comput. - UbiComp '15*. New York, New York, USA: ACM Press, sep 2015, pp. 403–407. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2750858.2804270

[42] I. S. MacKenzie and I. Scott, "Fitts' Law as a Research and Design Tool in Human-Computer Interaction," *Human?Computer Interact.*, vol. 7, no. 1, pp. 91–139, mar 1992. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1207/s15327051hci0701_3

[43] P. M. Fitts, "The information capacity of the human motor system in controlling the amplitude of movement." *J. Exp. Psychol.*, vol. 47, no. 6, pp. 381–391, 1954. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/h0055392

[44] J. Accot and S. Zhai, "Beyond Fitts' law," in *Proc. SIGCHI Conf. Hum. factors Comput. Syst. - CHI '97*. New York, New York, USA: ACM Press, 1997, pp. 295–302. [Online]. Available: https://dl.acm.org/citation.cfm?doid=258549.258760

[45] X. Cao and S. Zhai, "Modeling human performance of pen stroke gestures," in *Proc. SIGCHI Conf. Hum. factors Comput. Syst. - CHI '07*. New York, New York, USA: ACM Press, 2007, p. 1495.

[46] R. W. Proctor and K.-P. L. Vu, *Stimulus-response compatibility principles: Data, theory, and application*. CRC press, 2006.

[47] C. Forlines, D. Wigdor, C. Shen, and R. Balakrishnan, "Direct-touch vs. mouse input for tabletop displays," in *Proc. SIGCHI Conf. Hum. factors Comput. Syst. - CHI '07*. New York, New York, USA: ACM Press, apr 2007, p. 647. [Online]. Available: https://dl.acm.org/citation.cfm?doid=1240624.1240726

[48] C. B. Gibbs, "Controller design: Interactions of controlling limbs, time-lags and gain in positional and velocity systems," *Ergonomics*, vol. 5, no. 2, pp. 385–402, apr 1962. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/00140136208930602

[49] G. Casiez, D. Vogel, R. Balakrishnan, and A. Cockburn, "The Impact of Control-Display Gain on User Performance in Pointing Tasks," *Human-Computer Interact.*, vol. 23, no. 3, pp. 215–250, jul 2008. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/07370020802278163

[50] M. Nancel, D. Vogel, and E. Lank, "Clutching Is Not (Necessarily) the Enemy," in *Proc. 33rd Annu. ACM Conf. Hum. Factors Comput. Syst. - CHI '15*. New York, New York, USA: ACM Press, 2015, pp. 4199–4202. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2702123.2702134

[51] S. K. Card, J. D. Mackinlay, and G. G. Robertson, "A morphological analysis of the design space of input devices," *ACM Trans. Inf. Syst.*, vol. 9, no. 2, pp. 99–122, apr 1991. [Online]. Available: http://dl.acm.org/citation.cfm?id=123078.128726

[52] J. Accot and S. Zhai, "Scale effects in steering law tasks," *Proc. SIGCHI Conf. Hum. factors Comput. Syst. - CHI '01*, pp. 1–8, 2001. [Online]. Available: https://dl.acm.org/citation.cfm?doid=365024.365027

[53] Y. Guiard, “Difficulty and scale as the basic dimensions of aimed movement.” in *Proc. Tenth Int. Conf. Percept. Action*, University of Edinburgh, Scotland, 1999, p. p. 87.

[54] J. Williamson, “Continuous uncertain interaction,” 2006. [Online]. Available: http://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.433164

[55] R. A. Schmidt and T. D. Lee, *Motor control and learning: A behavioral emphasis, 4th ed.* Champaign, IL, US: Human Kinetics, 2005.

[56] F. Bevilacqua, B. Zamborlin, A. Sypniewski, N. Schnell, F. Guédy, and N. Rasamimanana, “Continuous realtime gesture following and recognition,” in *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 5934 LNAI. Springer, Berlin, Heidelberg, 2009, pp. 73–84. [Online]. Available: http://link.springer.com/10.1007/978-3-642-12553-9_7

[57] K. P. Murphy, *Machine Learning a probabilistic perspective*. MIT Press, 2012. [Online]. Available: https://mitpress.mit.edu/books/machine-learning-1

[58] A. J. Ijspeert, J. Nakanishi, H. Hoffmann, P. Pastor, and S. Schaal, “Dynamical movement primitives: Learning attractor models for motor behaviors,” pp. 328–373, feb 2013. [Online]. Available: https://www.mitpressjournals.org/doi/abs/10.1162/NECO_a_00393

[59] A. Oulasvirta, T. Roos, A. Modig, and L. Leppänen, “Information capacity of full-body movements,” in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst. - CHI ’13*. New York, New York, USA: ACM Press, apr 2013, p. 1289. [Online]. Available: http://dl.acm.org/citation.cfm?id=2470654.2466169

[60] M. Bachynskyi, G. Palmas, A. Oulasvirta, J. Steimle, and T. Weinkauf, “Performance and Ergonomics of Touch Surfaces: A Comparative Study using Biomechanical Simulation,” in *Proc. ACM CHI’15 Conf. Hum. Factors Comput. Syst.*, vol. 1. New York, New York, USA: ACM Press, apr 2015, pp. 1817–1826. [Online]. Available: http://dx.doi.org/10.1145/2702123.2702607

[61] M. Bachynskyi, G. Palmas, and T. Weinkauf, “Informing the Design of Novel Input Methods with Muscle Coactivation Clustering,” *Trans. Comput. Interact.*, vol. 21, no. 6, pp. 1–25, jan 2015. [Online]. Available: http://dx.doi.org/10.1145/2687921

[62] A. Oulasvirta, P. O. Kristensson, X. Bi, and A. Howes, *Computational interaction*.

[63] M. McGregor, B. Brown, and D. McMillan, “100 days of iPhone use,” in *Proc. Ext. Abstr. 32nd Annu. ACM Conf. Hum. factors Comput. Syst. - CHI EA ’14*. New York, New York, USA: ACM Press, sep 2014, pp. 2335–2340. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2559206.2581296

[64] Z. Yang, C. Yu, X. Yi, and Y. Shi, "Investigating Gesture Typing for Indirect Touch," *Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol.*, vol. 3, no. 3, pp. 1–22, sep 2019. [Online]. Available: http://dl.acm.org/citation.cfm?doid=3361560.3351275

[65] C. Holz and P. Baudisch, "Understanding touch," in *Proc. 2011 Annu. Conf. Hum. factors Comput. Syst. - CHI '11*. New York, New York, USA: ACM Press, may 2011, pp. 2501–2510. [Online]. Available: http://dl.acm.org/citation.cfm?doid=1978942.1979308

[66] A. Markussen, M. R. Jakobsen, and K. Hornbæk, "Vulture: a mid-air word-gesture keyboard," *Proc. 32nd Annu. ACM Conf. Hum. factors Comput. Syst. - CHI '14*, pp. 1073–1082, 2014. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2556288.2556964

[67] K. Vertanen, H. Memmi, J. Emge, S. Reyal, and P. O. Kristensson, "VelociTap," in *Proc. 33rd Annu. ACM Conf. Hum. Factors Comput. Syst. - CHI '15*. New York, New York, USA: ACM Press, 2015, pp. 659–668. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2702123.2702135

[68] B. Furht, *Handbook of augmented reality*. Springer Science & Business Media, 2011.

[69] S.-E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh, "Convolutional Pose Machines," in *IEEE Conf. Comput. Vis. Pattern Recognit.*, jun 2016, pp. 4724–4732. [Online]. Available: https://www.cv-foundation.org/openaccess/content_cvpr_2016/html/Wei_Convolutional_Pose_Machines_CVPR_2016_paper.html

[70] D. Herrera C., J. Kannala, and J. Heikkila, "Joint Depth and Color Camera Calibration with Distortion Correction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 10, pp. 2058–2064, oct 2012. [Online]. Available: http://ieeexplore.ieee.org/document/6205765/

[71] M. Firman, "RGBD Datasets: Past, Present and Future," in *IEEE Conf. Comput. Vis. Pattern Recognit. Work.*, jun 2016.

[72] F. Chollet and Others, "Keras," 2015. [Online]. Available: https://github.com/fchollet/keras

[73] T. Tieleman and G. Hinton, "Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude," *COURSERA Neural networks Mach. Learn.*, vol. 4, no. 2, pp. 26–31, 2012.

[74] W. Buxton, "A three-state model of graphical input," in *Human-computer Interact.*, vol. 90, 1990, pp. 449–456.

[75] P. Quinn and S. Zhai, "Modeling Gesture-Typing Movements," *Human–Computer Interact.*, pp. 1–47, aug 2016. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/07370024.2016.1215922

[76] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," 1988, pp. 139–183. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0166411508623869

[77] W. H. Press and S. A. Teukolsky, "Savitzky-Golay smoothing filters," *Comput. Phys.*, vol. 4, no. 6, pp. 669–672, 1990.

[78] J. Müller, A. Oulasvirta, and R. Murray-Smith, "Control Theoretic Models of Pointing," *ACM Trans. Comput. Interact.*, vol. 24, no. 4, pp. 1–36, aug 2017. [Online]. Available: http://dl.acm.org/citation.cfm?doid=3132166.3121431

[79] J. Deber, R. Jota, C. Forlines, and D. Wigdor, "How Much Faster is Fast Enough?" in *Proc. 33rd Annu. ACM Conf. Hum. Factors Comput. Syst. - CHI '15*. New York, New York, USA: ACM Press, 2015, pp. 1827–1836. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2702123.2702300

[80] N. Henze, S. Mayer, H. V. Le, and V. Schwind, "Improving software-reduced touchscreen latency," in *Proc. 19th Int. Conf. Human-Computer Interact. with Mob. Devices Serv. - MobileHCI '17*. New York, New York, USA: ACM Press, 2017, pp. 1–8. [Online]. Available: http://dl.acm.org/citation.cfm?doid=3098279.3122150

[81] I. J. Hubbard, M. W. Parsons, C. Neilson, and L. M. Carey, "Task-specific training: evidence for and translation to clinical practice," *Occup. Ther. Int.*, vol. 16, no. 3-4, pp. 175–189, sep 2009. [Online]. Available: http://doi.wiley.com/10.1002/oti.275

[82] S. Deterding, M. Sicart, L. Nacke, K. O'Hara, and D. Dixon, "Gamification. using game-design elements in non-gaming contexts," in *Proc. 2011 Annu. Conf. Ext. Abstr. Hum. factors Comput. Syst. - CHI EA '11*. New York, New York, USA: ACM Press, 2011, p. 2425. [Online]. Available: https://dl.acm.org/citation.cfm?doid=1979742.1979575

[83] J. Hamari, J. Koivisto, and H. Sarsa, "Does Gamification Work? – A Literature Review of Empirical Studies on Gamification," in *2014 47th Hawaii Int. Conf. Syst. Sci.* IEEE, jan 2014, pp. 3025–3034. [Online]. Available: http://ieeexplore.ieee.org/document/6758978/

[84] K. Seaborn and D. I. Fels, "Gamification in theory and action: A survey," *Int. J. Hum. Comput. Stud.*, vol. 74, pp. 14–31, feb 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1071581914001256

[85] H. W. Giessen, "Serious Games Effects: An Overview," *Procedia - Soc. Behav. Sci.*, vol. 174, pp. 2240–2244, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877042815009337

[86] S. Nicholson, "A RECIPE for Meaningful Gamification," in *Gamification Educ. Bus.* Cham: Springer International Publishing, 2015, pp. 1–20. [Online]. Available: http://link.springer.com/10.1007/978-3-319-10208-5_1

[87] "Exercise my game: Turning off-the-shelf games into exergames," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8215 LNCS, pp. 126–131, 2013.

[88] M. Ketcheson, L. Walker, and T. N. Graham, "Thighrim and Calf-Life," in *Proc. 2016 CHI Conf. Hum. Factors Comput. Syst. - CHI '16.* New York, New York, USA: ACM Press, 2016, pp. 2681–2692. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2858036.2858406

[89] G. A. V. Borg and B. J. Noble, "Perceived Exertion," *Exerc. Sport Sci. Rev.*, vol. 2, no. 1, 1974. [Online]. Available: https://journals.lww.com/acsm-essr/Fulltext/1974/00020/Perceived_Exertion.6.aspx

[90] "Intrinsic Motivation Inventory (IMI)." [Online]. Available: http://www.selfdeterminationtheory.org/intrinsic-motivation-inventory/

[91] J. M. Sietsema, D. L. Nelson, R. M. Mulder, D. Mervau-Scheidel, and B. E. White, "The use of a game to promote arm reach in persons with traumatic brain injury." *Am. J. Occup. Ther. Off. Publ. Am. Occup. Ther. Assoc.*, vol. 47, no. 1, pp. 19–24, jan 1993. [Online]. Available: https://ajot.aota.org/article.aspx?articleid=1872935

[92] X. Bao, Y. R. Mao, Q. Lin, Y. H. Qiu, S. Z. Chen, L. Li, R. S. Cates, S. F. Zhou, and D. F. Huang, "Mechanism of Kinect-based virtual reality training for motor functional recovery of upper limbs after subacute stroke," *Neural Regen. Res.*, vol. 8, no. 31, pp. 2904–2913, nov 2013. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/25206611

[93] M. Khademi, H. Mousavi Hondori, A. McKenzie, L. Dodakian, C. V. Lopes, and S. C. Cramer, "Free-hand interaction with leap motion controller for stroke rehabilitation," in *Proc. Ext. Abstr. 32nd Annu. ACM Conf. Hum. factors Comput. Syst. - CHI EA '14.* New York, New York, USA: ACM Press, 2014, pp. 1663–1668. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2559206.2581203

[94] N. Barrett, I. Swain, C. Gatzidis, and C. Mecheraoui, "The use and effect of video game design theory in the creation of game-based systems for

upper limb stroke rehabilitation," *J. Rehabil. Assist. Technol. Eng.*, vol. 3, p. 205566831664364, jun 2016. [Online]. Available: http://journals.sagepub.com/doi/10.1177/2055668316643644

[95] "Labyrinth (marble game)," nov 2018. [Online]. Available: https://en.wikipedia.org/wiki/Labyrinth_(marble_game)

[96] "Wire loop game," sep 2018. [Online]. Available: https://en.wikipedia.org/wiki/Wire_loop_game

[97] Konami, "Frogger," Game [Arcade], Tokyo, Japan, jun 1981.

[98] Namco, "Pac-Man," Game [Arcade], Tokyo, Japan, may 1980.

[99] J. E. Muñoz, M. Cameirão, S. Bermúdez i Badia, and E. R. Gouveia, "Closing the Loop in Exergaming - Health Benefits of Biocybernetic Adaptation in Senior Adults," in *Annu. Symp. Comput. Interact. Play Ext. Abstr. - CHI Play '18*. New York, New York, USA: ACM Press, 2018, pp. 329–339. [Online]. Available: http://dl.acm.org/citation.cfm?doid=3242671.3242673

[100] C. Plaisant, A. Druin, C. Lathan, K. Dakhane, K. Edwards, J. M. Vice, and J. Montemayor, "A Storytelling Robot for Pediatric Rehabilitation," oct 2000. [Online]. Available: https://drum.lib.umd.edu/handle/1903/1102

[101] A. K. Przybylski, C. S. Rigby, and R. M. Ryan, "A Motivational Model of Video Game Engagement," *Rev. Gen. Psychol.*, vol. 14, no. 2, pp. 154–166, 2010.

[102] M. Csikszentmihalyi, "Toward a Psychology of Optimal Experience," in *Flow Found. Posit. Psychol.* Dordrecht: Springer Netherlands, 2014, pp. 209–226. [Online]. Available: http://link.springer.com/10.1007/978-94-017-9088-8_14

[103] A. M. Limperos, M. G. Schmierbach, A. D. Kegerise, and F. E. Dardis, "Gaming Across Different Consoles: Exploring the Influence of Control Scheme on Game-Player Enjoyment," *Cyberpsychology, Behav. Soc. Netw.*, vol. 14, no. 6, pp. 345–350, jun 2011. [Online]. Available: http://www.liebertpub.com/doi/10.1089/cyber.2010.0146

[104] P. Sweetser and P. Wyeth, "GameFlow," *Comput. Entertain.*, vol. 3, no. 3, p. 3, jul 2005. [Online]. Available: https://dl.acm.org/citation.cfm?doid=1077246.1077253

[105] W. T. Powers and W. T. Powers, *Behavior: The control of perception*. Aldine Chicago, 1973.

[106] D. Trendafilov and R. Murray-Smith, "Information-Theoretic Characterization of Uncertainty in Manual Control," in *2013 IEEE Int. Conf. Syst. Man, Cybern.* IEEE, oct 2013, pp. 4913–4918. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6722590

[107] J. Williamson and R. Murray-Smith, "Rewarding the original: explorations in joint user-sensor motion spaces," in *Proc. 2012 ACM Annu. Conf. Hum. Factors Comput. Syst. - CHI '12*. New York, New York, USA: ACM Press, may 2012, p. 1717. [Online]. Available: http://dx.doi.org/10.1145/2207676.2208301

[108] E. Berdahl, M. Blandino, D. Baker, and D. Shanahan, "An Approach for Using Information Theory to Investigate Continuous Control of Analog Sensors by Humans," in *Proc. Audio Most. 2016 - AM '16*. New York, New York, USA: ACM Press, 2016, pp. 85–90. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2986416.2986450

[109] M. Zanfir, M. Leordeanu, and C. Sminchisescu, "The Moving Pose: An Efficient 3D Kinematics Descriptor for Low-Latency Action Recognition and Detection," pp. 2752–2759, 2013. [Online]. Available: http://openaccess.thecvf.com/content_iccv_2013/html/Zanfir_The_Moving_Pose_2013_ICCV_paper.html

[110] N. J. Cooke, "Varieties of knowledge elicitation techniques," *Int. J. Hum. Comput. Stud.*, vol. 41, no. 6, pp. 801–849, dec 1994. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S1071581984710834

[111] S. J. Russell and P. Norvig, *Artificial intelligence: a modern approach*. Malaysia; Pearson Education Limited,, 2016.

[112] C. M. Lu and N. J. Ferrier, "Repetitive Motion Analysis: Segmentation and Event Classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 2, pp. 258–263, 2004.

[113] D. Morris, T. S. Saponas, A. Guillory, I. Kelner, D. Morris, T. S. Saponas, A. Guillory, and I. Kelner, "RecoFit," in *Proc. 32nd Annu. ACM Conf. Hum. factors Comput. Syst. - CHI '14*. New York, New York, USA: ACM Press, 2014, pp. 3225–3234. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2556288.2557116

[114] B. Kruger, A. Vogele, T. Willig, A. Yao, R. Klein, and A. Weber, "Efficient Unsupervised Temporal Segmentation of Motion Data," *IEEE Trans. Multimed.*, vol. 19, no. 4, pp. 797–812, apr 2017. [Online]. Available: http://ieeexplore.ieee.org/document/7763809/

[115] O. Levy and L. Wolf, "Live Repetition Counting," pp. 3020–3028, 2015. [Online]. Available: https://www.cv-foundation.org/openaccess/content_iccv_2015/html/Levy_Live_Repetition_Counting_ICCV_2015_paper.html

[116] K. Perlin, Ken, Perlin, and Ken, "Improving noise," in *Proc. 29th Annu. Conf. Comput. Graph. Interact. Tech. - SIGGRAPH '02*, vol. 21, no. 3. New York, New York, USA: ACM Press, 2002, p. 681. [Online]. Available: https://dl.acm.org/citation.cfm?doid=566570.566636

[117] S. Bai, J. Z. Kolter, and V. Koltun, "An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling," mar 2018. [Online]. Available: http://arxiv.org/abs/1803.01271

[118] Y. Guiard, "On Fitts's and Hooke's laws: Simple harmonic movement in upper-limb cyclical aiming," *Acta Psychol. (Amst).*, vol. 82, no. 1-3, pp. 139–159, mar 1993. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/000169189390009G

[119] J. Gilliot, G. Casiez, and N. Roussel, "Impact of form factors and input conditions on absolute indirect-touch pointing tasks," in *Proc. 32nd Annu. ACM Conf. Hum. factors Comput. Syst. - CHI '14*. New York, New York, USA: ACM Press, 2014, pp. 723–732. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2556288.2556997

[120] B. Walther-franks, D. Wenig, J. Smeddinck, and R. Malaka, "Sportal : A First-Person Videogame Turned Exergame," *Proc. Mensch und Comput.*, pp. 539–542, 2013. [Online]. Available: https://dl.gi.de/handle/20.500.12116/7665

[121] M. Hofmann, K. Williams, T. Kaplan, S. Valencia, G. Hann, S. E. Hudson, J. Mankoff, and P. Carrington, ""Occupational Therapy is Making"," in *Proc. 2019 CHI Conf. Hum. Factors Comput. Syst. - CHI '19*. New York, New York, USA: ACM Press, 2019, pp. 1–13. [Online]. Available: http://dl.acm.org/citation.cfm?doid=3290605.3300544

[122] V. Levesque, L. Oram, K. MacLean, A. Cockburn, N. D. Marchuk, D. Johnson, J. E. Colgate, and M. A. Peshkin, "Enhancing physicality in touch interaction with programmable friction," in *Proc. 2011 Annu. Conf. Hum. factors Comput. Syst. - CHI '11*. New York, New York, USA: ACM Press, may 2011, p. 2481. [Online]. Available: http://dl.acm.org/citation.cfm?id=1978942.1979306

[123] D. Leithinger and H. Ishii, "Relief," in *Proc. fourth Int. Conf. Tangible, Embed. embodied Interact. - TEI '10*. New York, New York, USA: ACM Press, 2010, p. 221. [Online]. Available: https://dl.acm.org/citation.cfm?doid=1709886.1709928

Appendix A

Additional Pictures of Tools Created by Occupational Therapists

The initial meeting with an occupational therapist organised at the QEU hospital provided an occasion to photograh the various devices used to mitigate the motor limitations of the hands of her patients. Most contraptions were custom-made and tailored to the individual needs and physical properties of the patients, using a material that could be deformed once heated. The resulting shapes were then fixed in place after cooling.

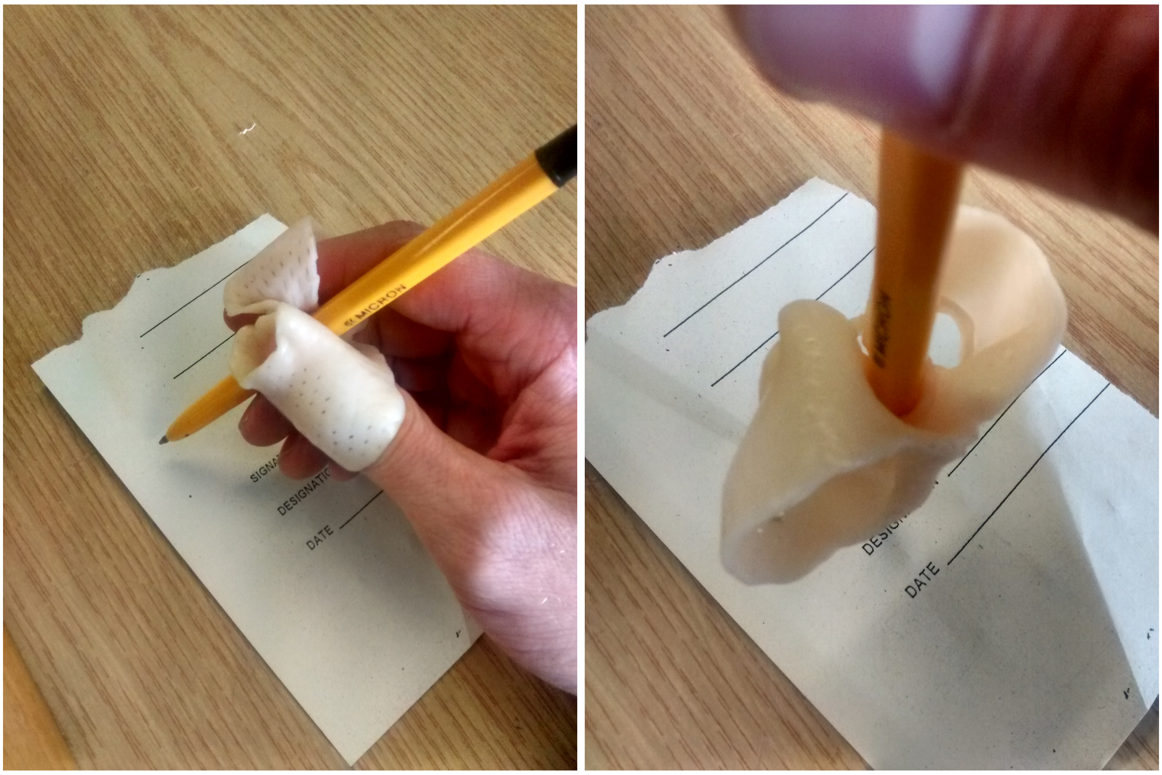

Figure A.1: Contraptions used for holding a pencil without requiring forces to be applied by the fingers. A special extrusion is made in the centre for the insertion of a pen. Note that the lack of flexibility of the material does not allow for different pencils to be used interchangeably.

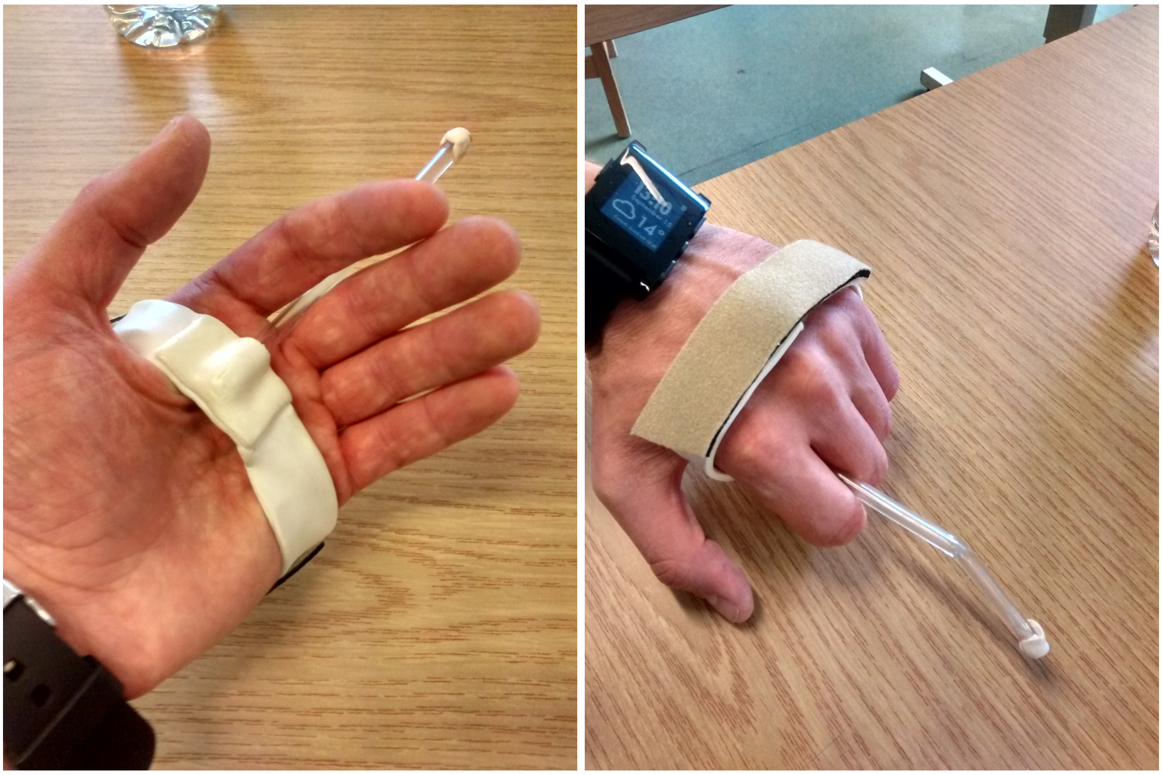

Figure A.2: Contraptions used to enable interactions with a touch sensitive surface, such as the one proposed by a smartphone or a tablet, without requiring extension of a single finger. Many patients presented some symptoms of rigidity in their forearms muscles with a consequence of a retraction of the fingers.

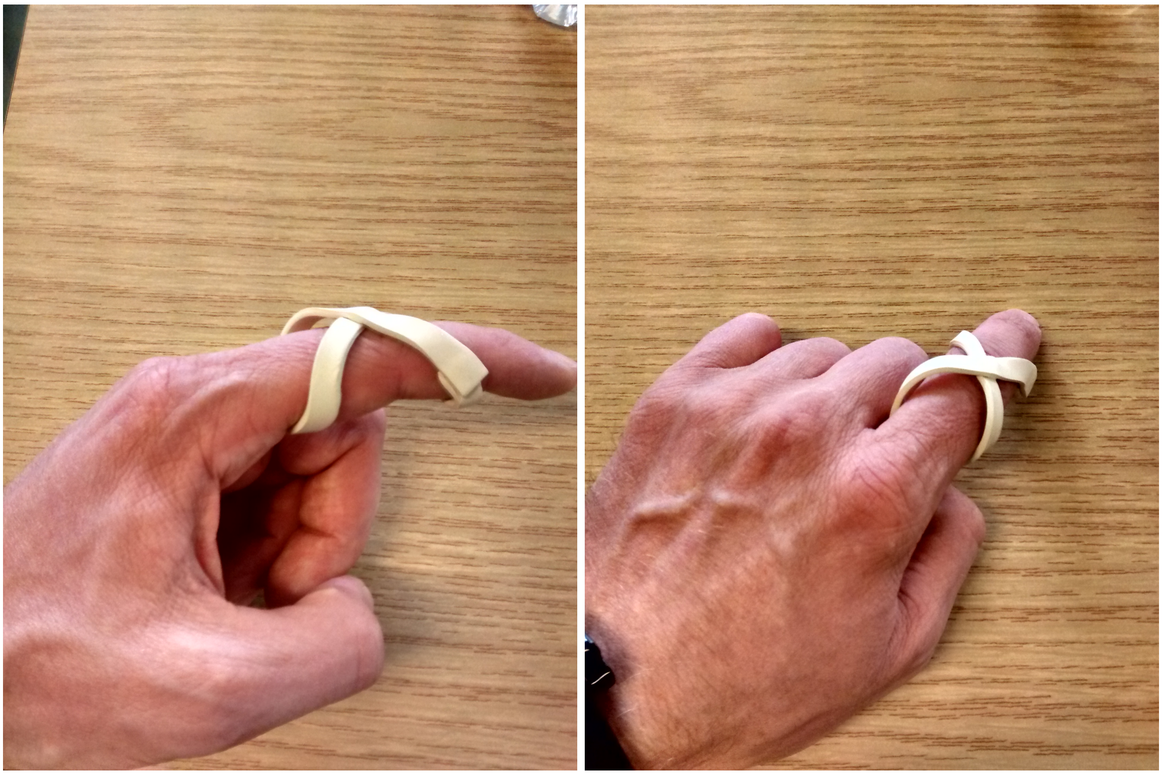

Figure A.3: Contraptions used to extend a specific finger, worn like a ring. This is meant to enable the interaction with a touch sensitive surface as well. Other mitigating strategies included the interaction with different parts of the hands.